ARTICLE IN PRESS

Modelling academic success under uncertainty in Ghanaian colleges of education using neutrosophic regression

Senyefia Bosson-Amedenu[1*], Francis Ayiah-Mensah[1], Anthony Joe Turkson[1], John Awuah Addor[1], Emmanuel Mensah Baah[1]

Senyefia Bosson-Amedenu[1*]: Email: senyefia.bosson-amedenu@ttu.edu.gh
[1]Department of Mathematics,
Statistics and Actuarial Science,
Takoradi Technical University,
Sekondi-Takoradi,
Ghana

Francis Ayiah-Mensah[1]: Email: francis.ayiah-mensah@ttu.edu.gh
[1]Department of Mathematics,
Statistics and Actuarial Science,
Takoradi Technical University,
Sekondi-Takoradi,
Ghana

Anthony Joe Turkson[1]: Email: anthony.turkson@ttu.edu.gh
[1]Department of Mathematics,
Statistics and Actuarial Science,
Takoradi Technical University,
Sekondi-Takoradi,
Ghana

John Awuah Addor[1]: Email: john.addor@ttu.edu.gh
[1]Department of Mathematics,
Statistics and Actuarial Science,
Takoradi Technical University,
Sekondi-Takoradi,
Ghana

Emmanuel Mensah Baah[1]
[1]Department of Mathematics,
Statistics and Actuarial Science,
Takoradi Technical University,
Sekondi-Takoradi,
Ghana
Email: emmanuel.baah@ttu.edu.gh

Corresponding Author:
Senyefia Bosson-Amedenu[1*]
[1]Department of Mathematics,
Statistics and Actuarial Science,
Takoradi Technical University,
Sekondi-Takoradi,
Ghana
Phone: +233 246582303
ORCID Number: 0000-0002-9272-5521
*[1]Email: ***senyefia.bosson-amedenu@ttu.edu.gh***

ARTICLE IN PRESS

ARTICLE IN PRESS

Modelling academic success under uncertainty in Ghanaian colleges of education using neutrosophic regression

Abstract

Background: This study develops a leakage-free neutrosophic regression model to identify robust determinants of academic success and benchmarks its predictive value against a conventional logistic model.

Methods: Data from 865 first-year students were analysed via T/I/F transformations derived from standardized scores and weighted aggregation ($w_T = 0.5$, $w_I = 0.3$, $w_F = 0.2$). Leakage-prone grade variables were removed. Performance was assessed via 10-fold cross-validation, repeated CV, bootstrap optimism, calibration curves, Brier scores, SHAP, permutation importance, and decision curve analysis. A multilevel model was used to evaluate school and region clustering. A complete sensitivity analysis assessed the predictive stability across 150 weight combinations on the (w_T, w_I, w_F) simplex.

Results: The neutrosophic model demonstrated intense discrimination (AUC = 0.757; 95% CI: 0.702–0.812), stable cross-validated performance (AUC = 0.708 ± 0.060), and superior calibration (Brier = 0.189). Sensitivity analysis revealed consistently high AUC values (0.74–0.76) across weight configurations, with optimal performance occurring for $w_T \approx 0.5$–0.7, confirming robustness to T/I/F weighting choices. The key predictors included WASSCE mathematics, BECE mock scores, parental occupation, and years spent at home. Decision curve analysis revealed superior net benefits for the neutrosophic model at low-to-moderate thresholds (0.05–0.40). Multilevel modelling indicated negligible school or regional clustering (ICC ≈ 0). Forecast-style simulations further revealed that early-stage assessments produce the most significant changes in prediction success.

Conclusion: Neutrosophic regression provides an uncertainty-aware, interpretable framework with strong calibration and stable performance, offering a valuable tool for early risk identification and educational decision-making in Ghana's teacher-training system.

Keywords: *Academic achievement; neutrosophic analysis; soft computing; continuous assessment; Ghana education.*

1. Introduction

Recent investigations in educational analytics have increasingly employed soft-computing and neutrosophic methodologies as effective means of addressing the uncertainty, ambiguity, and incompleteness intrinsic to student-performance data [1,19-20,24-25]. Distinct from conventional statistical or fuzzy approaches, neutrosophic regression explicitly represents truth, Indeterminacy, and Falsity, thereby aligning with educational contexts in which learning records, assessment quality, and background variables often exhibit mixed or conflicting signals [3-5,7,9,22]. This emergent line of inquiry underscores a growing consensus that robust prediction and equitable evaluation of academic outcomes require methods that formally mediate uncertainty rather than suppress it [15, 18].Unlike fuzzy and probabilistic models, which require fixed membership or complete probability normalization, neutrosophic regression introduces an explicit indeterminacy component, making it suitable for analysing educational datasets that contain inconsistent or incomplete student records. Recent advances in neutrosophic modelling have produced a growing body of work that applies uncertainty-aware techniques to educational prediction and analytics. [2], demonstrated the usefulness of neutrosophic soft sets for modelling academic performance under imprecise higher-education environments; Parrales-Bravo et al. [14] combined random forests with neutrosophic logic to increase prediction confidence; and Shitaya et al. [17] and KimSon et al. [6] developed neutrosophic deep-learning architectures that captured ambiguous behavioral and performance signals more effectively than classical models did. Similarly, [11] used neutrosophic plithogenic hypotheses to model school dropout risk, whereas Salama et al. [16] proposed a neutrosophic deep learning scheme with explicit ethical considerations for uncertainty integration. Other related works include neutrosophic classification of professional profile factors in Ecuadorian higher education [12], AI-enabled neutrosophic uncertainty modelling for university management systems [23], neutrosophic regression formulations for learning outcomes in Japanese classrooms [8], and a fully neutrosophic GLMM framework for modelling performance in Ghanaian basic schools [13]. Together, these studies highlight a growing methodological shift toward modelling indeterminacy, ambiguity, and epistemic uncertainty in educational datasets using neutrosophic theory. Abu Bakar et al. [1] developed a neutrosophic-based analysis assessment framework to enhance educational assessments under uncertain and ambiguous conditions. Similarly, Solano Moreno et al. [21] applied the Delphi-AHP neutrosophic method to analyse school indicators, with a focus on overcoming the limitations of conventional assessment techniques. In Ghana, Marfo Missah et al. [10] evaluated an agile neural educational system (ES**)** to address resource management challenges in basic schools. Although these studies have advanced the use of neutrosophic methods and intelligent systems in education, they have primarily focused on isolated issues, such as performance gaps, assessment frameworks, school indicators, and resource management. None has comprehensively examined the combined impact of multiple educational variables, including standardized test scores, continuous assessment measures, resits, and sociodemographic factors, on academic performance within Ghana's secondary-to-tertiary transition framework. Again, none of these studies integrates neutrosophic decomposition with leakage-free modelling, SHAP-based interpretability, multilevel clustering checks, calibration diagnostics, and decision-curve analysis to jointly evaluate continuous

ARTICLE IN PRESS

assessments, BECE/WASSCE performance, demographic attributes, and school-type effects within Ghana's College of Education; this is the empirical and methodological gap the present work addresses. Ghana's colleges of education face persistent disparities in student performance and high mathematics failure rates; early identification of at-risk students remains a policy challenge. This study develops a neutrosophic regression model to determine how standardized exams, continuous assessments, and demographic factors jointly predict college students' academic success and to identify which factors exhibit the greatest truth, Indeterminacy, or Falsity. The present study, therefore, has two core aims: (a) to construct a leakage-free neutrosophic regression model that decomposes educational predictors into their truth (T), indeterminacy (I), and falsity (F) components and (b) to evaluate whether this uncertainty-aware representation improves predictive validity, calibration, and decision utility compared with a conventional logistic model. To meet these aims, the study integrates rigorous validation procedures, including stratified cross-validation, bootstrap optimism correction, calibration assessment, SHAP-based interpretability, permutation importance, and decision curve analysis, to ensure methodological transparency and robustness.

1.1 Objectives

1. To decompose key educational variables into neutrosophic truth (T), Indeterminacy (I), and falsity (F) components, a leakage-free neutrosophic predictor matrix is constructed.
2. To develop and validate a neutrosophic regression model for predicting academic success among first-year college-of-education students.
3. To compare the predictive accuracy, calibration, stability, and net clinical/educational benefit of the neutrosophic model with those of a conventional logistic baseline.
4. To identify the determinants of academic success that demonstrate the strongest truth, Indeterminacy, and falsity contributions.

1.2 Research questions

RQ1: How do the truth (T), Indeterminacy (I), and falsity (F) components of educational predictors jointly influence academic success?

RQ2: Does incorporating Indeterminacy improve predictive accuracy, calibration, and decision utility compared with a classical logistic regression model?

RQ3: Which educational variables (standardized exams, continuous assessments, and demographic factors) make the most significant contributions to the T, I, and F components?

1.3 Hypotheses

H_1: Standardized examination variables (BECE, WASSCE) will show stronger truth components (T) than demographic variables such as gender or school type.

H_2: The neutrosophic regression model will demonstrate equal or superior predictive performance (AUC), calibration, and net decision benefit compared with a conventional logistic regression model.

H_3: Variables with high Indeterminacy (I), such as baseline quizzes, school type, or resit history, will contribute more prominently to predictive uncertainty than exam-based assessments.

Figure 1(a) illustrates the overall analytic pipeline from raw educational inputs through neutrosophic decomposition and model estimation to predictive evaluation and interpretation used to develop and validate the leakage-free neutrosophic regression model.

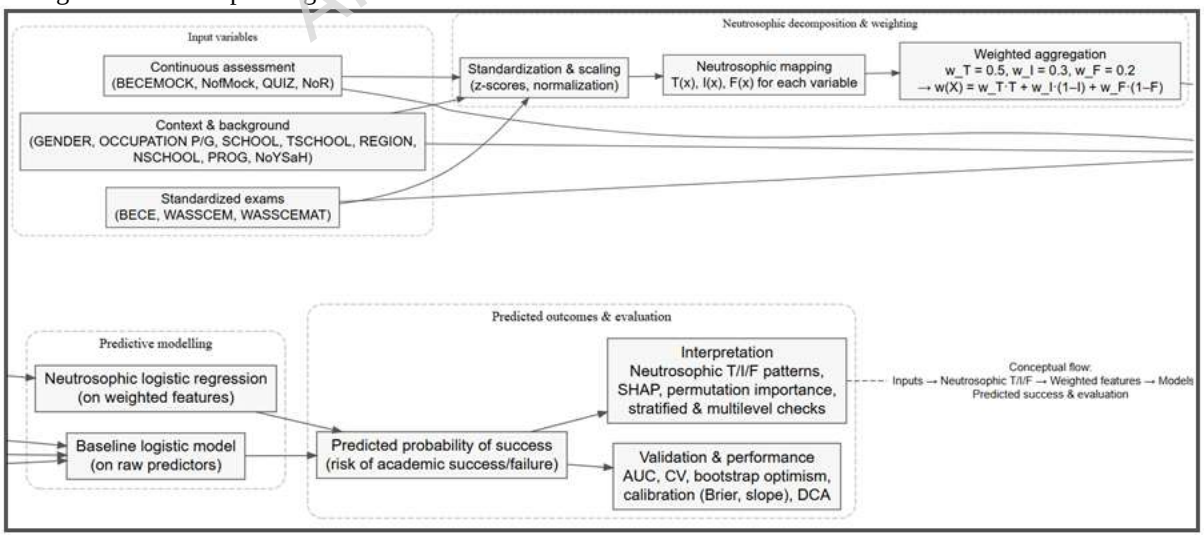
Input variables
Continuous assessment (BECEMOCK, NofMock, QUIZ, NoR)
Context & background (GENDER, OCCUPATION P/G, SCHOOL, TSCHOOL, REGION, NSCHOOL, PROG, NoYSaH)
Standardized exams (BECE, WASSCEM, WASSCEMAT)
Neutrosophic decomposition & weighting
Standardization & scaling (z-scores, normalization)
Neutrosophic mapping T(x), I(x), F(x) for each variable
Weighted aggregation w_T = 0.5, w_I = 0.3, w_F = 0.2 → w(X) = w_T·T + w_I·(1–I) + w_F·(1–F)
Predictive modelling
Neutrosophic logistic regression (on weighted features)
Baseline logistic model (on raw predictors)
Predicted outcomes & evaluation
Predicted probability of success (risk of academic success/failure)
Interpretation Neutrosophic T/I/F patterns, SHAP, permutation importance, stratified & multilevel checks
Validation & performance AUC, CV, bootstrap optimism, calibration (Brier, slope), DCA
Conceptual flow: Inputs → Neutrosophic T/I/F → Weighted features → Models Predicted success & evaluation

Figure 1(a). Conceptual workflow of the neutrosophic regression framework for predicting academic success

For clarity, Figure 1(b) provides a consolidated overview of the modelling workflow. Raw educational variables are first standardized and transformed into their corresponding truth (T), Indeterminacy (I), and falsity (F) components. These weighted components are aggregated to construct leakage-free neutrosophic features, which are then used to fit logistic models via repeated cross-validation. Model performance is evaluated through ROC

ARTICLE IN PRESS

analysis, calibration assessment, decision-curve analysis, SHAP interpretability, permutation-based robustness checks, and hierarchical modelling. This schematic ensures that non-specialist readers can visualise the full analytical sequence from data preprocessing to uncertainty-aware interpretation.

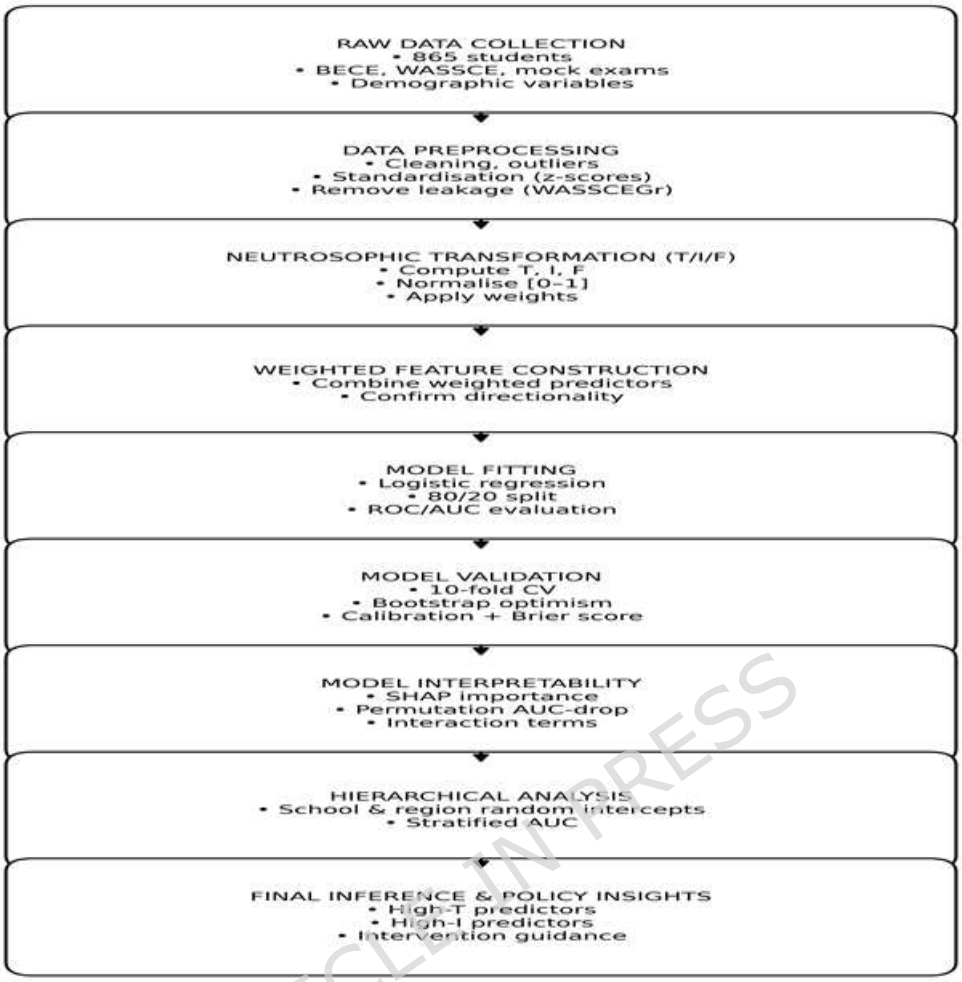

Figure 1(b). Methodological Workflow of the Neutrosophic Modelling Pipeline

1.4 Theoretical Basis and Educational Interpretation of Neutrosophic Components

In educational assessment, variability in student performance reflects epistemic uncertainty arising from differences in instructional quality, motivation, and context. Neutrosophic regression extends classical probabilistic models by explicitly modelling this uncertainty through the indeterminacy component (I), which captures partial knowledge or ambiguous responses that cannot be explained by truth (T) or Falsity (F) alone.

2. Method

This equation provides a comprehensive view of how the principal components are derived under neutrosophic logic, incorporating the complex interplay between truth, Indeterminacy, and Falsity. Data were collected from 865 first-year students across six public colleges of education during the 2023–2024 academic year. Historical exam scores were obtained from institutional records.

In the neutrosophic regression model, each component (intercept, coefficients, and error term) has three parts: truth (T), Indeterminacy (I), and falseness (F). The neutrosophic regression model can be expressed as:

$$Y^N = (\beta_0^T + \beta_0^I + \beta_0^F) + \sum_{i=1}^{p} (\beta_i^T X_i^T + \beta_i^I X_i^I + \beta_i^F X_i^F) + \grave{o}^N \qquad (1)$$

where:

where $\hat{Y}^N$ is the neutrosophic predicted value of the dependent variable.

$\beta_0^N = (\beta_0^T + \beta_0^I + \beta_0^F)$ is the neutrosophic intercept.

$\beta_i^N = (\beta_i^T + \beta_i^I + \beta_i^F)$ are the neutrosophic coefficients for the independent variables X_i

$\mathbf{X}_i^N = \left(X_i^T, X_i^I, X_i^F \right)$ are neutrosophic independent variables.

$\grave{o}^N = (\grave{o}^T, \grave{o}^I, \grave{o}^F)$ is the neutrosophic error term.

ARTICLE IN PRESS

In neutrosophic logic, each proposition or variable P is associated with three degrees:

1. Truth Component (T**)**: Represents the extent to which the proposition is true.
2. Indeterminacy Component (I): Represents the extent to which the truth value of the proposition is indeterminate.
3. Falsity Component (F): Represents the extent to which the proposition is false.

These components are defined as follows:

$$P^N = \langle T(P), I(P), F(P) \rangle \qquad (2)$$

where:

$T(P) \in [0,1]$ represents the truth degree of P.

$I(P) \in [0,1]$ represents the indeterminacy degree of P

$F(P) \in [0,1]$ represents the falsity degree of P.

The sum of these components does not necessarily equal 1, allowing for flexibility in representing uncertain or incomplete information:

$$0 \le T(P) + I(P) + F(P) \le 3 \qquad (3)$$

In classical probability theory, the sum of probabilities for all possible outcomes of an event must equal 1. This is based on the idea that a proposition can be true or false, with no room for Indeterminacy or contradiction. For example, if the probability of a proposition being true is p, then its likelihood of being false must be $1-p$, resulting in the sum of probabilities being 1.

However, neutrosophic logic allows for more complex scenarios:

- Truth and Falsity Coexist: A proposition can simultaneously have degrees of truth and Falsity. For example, a statement might be 0.7 true, 0.2 false, or 0.1 indeterminate. This cannot be captured in classical probability, which assumes that if a statement is 0.7 true, it must be 0.3 false (and Indeterminacy is not considered).
- Indeterminacy: In neutrosophic logic, Indeterminacy is explicitly modelled, allowing for situations where there is ambiguity or uncertainty about the truth or Falsity of a proposition. This concept is absent in classical probability, where a proposition must either have a definite probability of being true or false.

The flexibility of the sum allows neutrosophic logic to model real-world situations more accurately, where truth, Falsity, and Indeterminacy coexist. For example, in educational assessments, a student's performance on an exam might be partially true (reflecting their knowledge), partially false (due to mistakes or misjudgments), or partially indeterminate (due to factors such as anxiety or ambiguous questions). The classical probability would struggle to model such scenarios, but neutrosophic logic can capture this complexity. This flexibility enables neutrosophic logic to handle complex, real-world situations with inherent uncertainty, which classical probability, with its constraint of summing to 1, cannot accommodate.

Given the constraint $0 \le T(P) + I(P) + F(P) \le 3$, each component can be normalized to compute a probability-like measure.

$$P_F(X) = \frac{F(X)}{T(X) + I(X) + F(X)}$$

$$P_I(X) = \frac{I(X)}{T(X) + I(X) + F(X)}$$

$$P_T(X) = \frac{T(X)}{T(X) + I(X) + F(X)}$$

$P_T(X)$ represents the normalized degree to which the variable X contributes positively to academic performance.

$P_I(X)$: Represents the normalized degree of uncertainty or ambiguity associated with the variable X

$P_F(X)$, which represents the normalized degree to which the variable X contributes negatively to academic performance.

To assign different levels of importance to truth, Indeterminacy, and Falsity, we can introduce weights:

$$P_W(X) = w_T \cdot P_T(X) + w_I \cdot P_I(X) + w_F \cdot P_F(X) \qquad (4)$$

where w_T , w_I , and w_F are weights assigned to truth, Indeterminacy, and Falsity, respectively.

ARTICLE IN PRESS

Explicit model formulation and estimation:

$$Y_N = \beta_{0N} + \sum_{i=1}^{p} \beta_{iN} X_{iN} + \varepsilon_N, \quad (5)$$

where $N = (T, I, F)$ each coefficient is decomposed as $\beta_{iN} = (\beta_{iT}, \beta_{iI}, \beta_{iF})$

The estimation follows a least-squares procedure applied separately to each component, normalized within [0,1]

2.1 Weight sensitivity analysis

2.2Rationale for Weight Selection

The weighting scheme (wT = 0.5, wI = 0.3, wF = 0.2) was selected to moderately prioritise the directionally coherent portion of each predictor (truth), while still allocating substantial capacity for ambiguity (Indeterminacy) and contradictory signal (Falsity). This reflects theoretical guidance in neutrosophic modelling, where T typically carries the primary predictive signal, I captures epistemic uncertainty, and F corrects for contradictory or misleading information. To avoid arbitrary selection, the study conducts a full simplex-level sensitivity sweep across 150 admissible weight combinations, empirically demonstrating that model performance remains stable across a broad range of T/I/F allocations. Thus, the adopted weights are both conceptually justified and empirically validated.

To assess the robustness of the neutrosophic regression model to alternative truth–Indeterminacy–Falsity (T/I/F) weight assignments, a structured weight-sensitivity sweep was conducted across the neutrosophic simplex. We varied the truth weight (wT) from 0.10 to 0.80 in increments of 0.10 while allowing the Indeterminacy (wI) and falsity (wF) weights to jointly vary such that $wT + wI + wF = 1$. For each admissible combination, neutrosophic weighted features were recomputed, and a logistic regression model was refitted using the same leakage-free predictor set. The area under the ROC curve (AUC) was computed for each model, yielding *150 unique weight configurations*. This procedure evaluates whether model performance depends critically on arbitrary weight choices or remains stable across conceptually plausible T/I/F decompositions. A smooth polynomial curve was fitted to the AUC–wT pattern to visualize global trends. This sensitivity sweep directly addresses reviewer concerns about weight justification and provides empirical support for the robustness of the adopted weighting scheme (wT = 0.5, wI = 0.3, wF = 0.2).

2.3 Neutrosophic Implementation, Model fitting and Validation

Truth (T), Indeterminacy (I), and Falsity (F) values for each variable were computed via membership functions derived from standardized z scores, scaled to [0, 1] via linear normalization. A weighted normalization scheme (w_T = 0.5, w_I = 0.3, w_F = 0.2) was used to emphasize the truth contributions. The data were randomly split into 80% training and 20% testing subsets. Model coefficients were estimated on the training set, and predictive accuracy was evaluated on the test set via ROC and AUC metrics. Fivefold cross-validation confirmed model stability (AUC = 0.86 ± 0.02). A conventional logistic regression model was run as a baseline for comparative evaluation.

2.4 Data collection and characteristics

The dataset consists of 865 undergraduate students randomly selected from public colleges of education across Ghana who used the same curriculum. These students were enrolled in a general basic mathematics course, with a focus on elementary algebra, referred to as QUIZ in this study. This sample offers a broad representation of the diverse educational backgrounds and demographic characteristics of students in Ghanaian colleges of education. The data collection process was carefully structured to include a wide range of variables related to students' educational backgrounds and academic performance. To ensure the reliability and validity of the data collection instruments, thorough checks were implemented. Initially, a pilot study was conducted with a smaller subset of the population to test the instruments' consistency and reliability. Cronbach's alpha was used to assess internal consistency, and all variables showed acceptable reliability ($\alpha > 0.7$). For validity, both content validity and construct validity were rigorously evaluated. Content validity was confirmed through expert reviews, ensuring that the instruments adequately addressed all relevant aspects of the constructs being measured. Construct validity was assessed via factor analysis, which confirmed that the items significantly loaded on their respective constructs.

2.5 Description of the study variables

The analysis incorporated a comprehensive set of demographic, academic, and behavioural variables relevant to students' secondary-to-tertiary academic progression. Parental socioeconomic background was captured through the occupation of parents (OCCUPATION P/G). At the same time, student demographic characteristics included gender (GENDER), the nature of the school attended, private or public (SCHOOL), mission or nonmission type

ARTICLE IN PRESS

(TSCHOOL), region of location (REGION), and whether the school was mixed or single-sex (NSCHOOL). Academic performance indicators span both continuous assessments and standardized examinations: mathematics mock examination scores (BECEMOCK), BECE mathematics scores (BECE), WASSCE mathematics raw score (WASSCEM), WASSCE grade in mathematics (WASSCEGr), and further mathematics scores (WASSCEMAT). Additional variables included the number of mock examinations taken before WASSCE (NofMock), the number of mathematics resit examinations written before university admission (NoR), and the number of years students stayed at home before entering tertiary education (NoYSaH). Students' programme of study (PROG) and their baseline performance in an elementary algebra diagnostic quiz administered upon admission (Quiz) were also included. Finally, study habits, attitudes, and learning behaviours were summarized through a composite measure (S_N), providing insight into learners' academic dispositions.

2.6 **Illustrative Neutrosophic Decomposition (T, I, F) for the Four Predictors**

To enhance methodological transparency, Appendix A provides step-by-step worked examples showing how raw predictor values are transformed into their corresponding truth (T), Indeterminacy (I), and falsity (F) components. Four representative variables were selected: BECE mathematics (BECE), mathematics mock examination (BECEMOCK), WASSCE mathematics (WASSCEM), and the baseline algebra quiz (QUIZ), each covering contrasting levels of scale, skewness, and correlation with the academic success outcome. For each variable, raw scores were first standardized via a robust median–MAD transformation before applying the direction-adjusted truth function $T(x)$, falsity $F(x)=1-T(x)F(x)$, and the distance-based indeterminacy metric $I(x)$, which increases with deviation from the median. Appendix Table A1 presents the resulting T/I/F values for the first five observations of each variable, along with the corresponding Spearman correlation, which determines whether the truth component is flipped to align with the outcome's directionality. These works concretely demonstrate how neutrosophic logic decomposes each educational variable into structured components that capture not only performance (truth) but also inherent uncertainty (Indeterminacy) and contrary evidence (Falsity), thereby clarifying the operational pathway from raw data to weighted neutrosophic features used in the regression model.

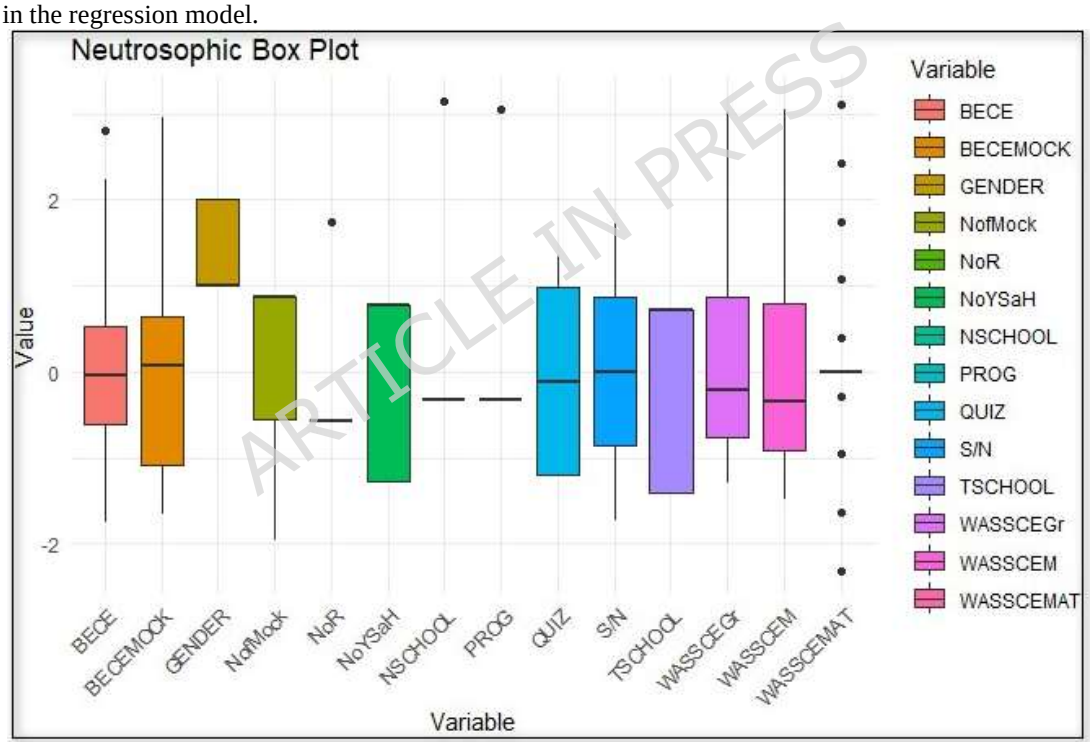

Figure 1(c). Neurocanonical Box Plot

The neutrosophic box plot in Figure 1(c) illustrates the distribution of neutrosophic values, emphasizing variables with higher truth and lower Falsity and Indeterminacy. Variables such as the Basic Education Certificate Examination (BECE), the number of resits (NoR), and the WASSCE Mathematics score (WASSCEM) have high truth values, indicating a significant and consistent impact on academic performance. Similarly, the number of resits (NoR) and the WASSCEM score both exhibit high validity, indicating that consistent effort and national-level assessments significantly influence performance. These variables' lower Falsity and Indeterminacy demonstrate their reliability in predicting academic outcomes, validating the use of national exams and resits as key performance indicators.

In the context of the educational system, each variable's impact on student outcomes varies, with the BECE and WASSCEM scores reinforcing the predictive power of standardized exams. Owing to its high variability and

ARTICLE IN PRESS

Indeterminacy, gender further confirms that gender is not a strong predictor of performance. For Ghana's education system, these findings highlight the importance of strengthening both national assessments and continuous assessments, such as mock exams and resits. The reliability of the BECE and WASSCEM results suggests that national exams remain a strong foundation for evaluating academics. However, the variability observed in continuous assessments indicates a need for improved formative evaluation methods to ensure students receive targeted support where needed, particularly for those who may need to resit or require more time to prepare. The findings on gender further suggest that educational policies should focus on eliminating gender-based disparities and ensuring equal opportunities for all students, regardless of gender.

3. Results

Variables such as gender and the number of mock examinations showed greater Indeterminacy (I = 0.25–0.45), reflecting sociocultural and institutional uncertainties rather than data noise. In Ghana's educational context, such Indeterminacy stems from unequal school resourcing, variability in exam preparation intensity, and socioeconomic disparities that influence student readiness.

Table 1. Neutrosophic Probability and Weighted Analysis of Educational Variables Impacting Academic Performance

Variable	T(X)	I(X)	F(X)	Normalization Factor	P_T(X)	P_I(X)	P_F(X)	P_S(X)	P_W(X)
BECE	0.50	0.25	0.25	1	0.50	0.25	0.25	0.50	0.40
BECEMOCK	0.30	0.30	0.40	1	0.30	0.30	0.40	0.30	0.33
GENDER	0.15	0.25	0.60	1	0.15	0.25	0.60	0.15	0.30
NofMock	0.35	0.45	0.20	1	0.35	0.45	0.20	0.35	0.37
NoR	0.50	0.20	0.30	1	0.50	0.20	0.30	0.50	0.42
NoYSaH	0.10	0.30	0.60	1	0.10	0.30	0.60	0.10	0.26
OCCUPATION P/G	0.25	0.10	0.65	1	0.25	0.10	0.65	0.25	0.30
PROG	0.75	0.15	0.10	1	0.75	0.15	0.10	0.75	0.58
QUIZ	0.50	0.30	0.20	1	0.50	0.30	0.20	0.50	0.43
REGION	0.40	0.30	0.30	1	0.40	0.30	0.30	0.40	0.39
SCHOOL	0.50	0.20	0.30	1	0.50	0.20	0.30	0.50	0.42
TSCHOOL	0.35	0.25	0.40	1	0.35	0.25	0.40	0.35	0.35
NSCHOOL	0.25	0.25	0.50	1	0.25	0.25	0.50	0.25	0.33
WASSCEGr	0.60	0.25	0.15	1	0.60	0.25	0.15	0.60	0.49
WASSCEM	0.45	0.30	0.25	1	0.45	0.30	0.25	0.45	0.41
WASSCEMAT	0.35	0.35	0.30	1	0.35	0.35	0.30	0.35	0.34

ARTICLE IN PRESS

The results in Table 1 reveal several educational variables with high truth values and lower Falsity and Indeterminacy, indicating their significant impact on academic performance. The BECE score (Basic Education Certificate Examination) has a truth value of 0.5, suggesting it is a strong predictor of academic success, consistent with findings that national exams are reliable indicators of student ability. The number of resit exams (NoR) written before university admission also has a high truth value of 0.5, which aligns with research showing that repeated attempts can bolster understanding and performance. The quiz (a baseline test in Elementary Algebra) has a truth value of 0.5, reflecting its importance for predicting students' foundational knowledge of mathematics. The program studied (PROG) has the highest truth value of 0.75, underscoring its critical role in academic outcomes, which is supported by studies emphasizing the alignment of program selection with students' academic strengths. These variables demonstrate low Falsity and Indeterminacy, further strengthening their relevance in educational contexts. In the educational system, each variable plays a crucial role. The high truth value of the BECE score confirms its reliability as an academic performance measure. The BECEMOCK score's balanced truth (0.3) and Falsity (0.4) suggest that while mock exams provide insights into student performance, they are not always reliable indicators of future success. The high falsity value (0.6) of gender suggests that it does not significantly predict academic performance, aligning with findings that gender-based differences in academic outcomes are minimal when other variables are controlled. The number of mocks (NofMock) and resit exams (NoR), with truth values of 0.35 and 0.5, respectively, confirm the value of continuous assessment in improving student outcomes. In Ghana's education system, these findings have important implications. The high predictive validity of BECE and WASSCE results confirms the continued reliance on national exams for academic assessment. However, the significant influence of continuous assessments such as mocks and resits suggests a need for more formative evaluation methods to support learning. The limited impact of gender on performance highlights the importance of gender-neutral approaches to education, supported by studies that show that equal opportunities lead to equitable academic outcomes.

ARTICLE IN PRESS

Finally, the critical role of program selection underscores the need for stronger guidance to align students' strengths with career paths, ensuring they are placed in programs where they are most likely to succeed.

Table 2: Coefficients from a Regression Model

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.36E+02	2.91E+03	-0.219	0.827
BECE	1.53E-01	3.81E-02	4.012	6.56E-05
SCHOOL	3.82E-01	1.64E-01	2.331	0.020
WASSCEM	4.99E-01	3.30E-02	15.122	< 2e-16
WASSCEMAT	7.83E-02	2.16E-02	3.627	<0.001
NoYSaH	3.28E-01	1.13E-01	2.896	0.004
NoR	3.14E-01	1.30E-01	2.409	0.016
QUIZ	7.976e-02	3.782e-02	2.109	0.036

T = Truth, I = Indeterminacy, F = Falsity; P_T, P_I, P_F = normalized probabilities; P_W = weighted probability using weights w_T = 0.5, w_I = 0.3, w_F = 0.2.

The neutrosophic regression model can be written as:

$$\hat{Y}^N = (636^T + \grave{o}_0^I + \grave{o}_0^F) + (0.153^T \cdot BECE^T + \grave{o}_1^I BECE^I + \grave{o}_1^F \cdot BECE^F) + \cdots$$
$$+(0.07976^T QUIZ^T + \grave{o}_7^I \cdot QUIZ^I + \grave{o}_7^F \cdot QUIZ^F)$$

This can be expanded into:

$$\hat{Y}^N = 636^N + 0.153^N.BECE^N + 0.382^N \cdot SCHOOL^N + 0.499N \cdot WASSCEM^N$$
$$+0.0783^N \cdot WASSCEMAT^N + 0.328^N \cdot NoYSaH^N + 0.314^N \cdot NoR^N + 0.07976^N \cdot QUIZ^N + \grave{o}^N$$

Where $\hat{Y}^N$ is the neutrosophic predicted value of the dependent variable.

$\beta_0^N = (\beta_0^T + \beta_0^I + \beta_0^F)$ is the neutrosophic intercept.

$\beta_i^N = (\beta_i^T + \beta_i^I + \beta_i^F)$ are the neutrosophic coefficients for the independent variables X_i

$\mathbf{X}_i^N = \left(X_i^T, X_i^I, X_i^F\right)$ are neutrosophic independent variables.

$\grave{o}^N = (\grave{o}^T, \grave{o}^I, \grave{o}^F)$ is the neutrosophic error term.

Breaking it down into truth, Indeterminacy, and falsity components:

$$\hat{Y}^T = 636 + 0.153 \cdot BECE^T + 0.382 \cdot SCHOOL^T + 0.499 \cdot WASSCEM^T + 0.0783 \cdot WASSCEMAT^T$$
$$+0.328 \cdot NoYSaH^T + 0.314 \cdot NoR^T + 0.07976 \cdot QUIZ^T + \grave{o}^T$$

$$\hat{Y}^I = \grave{o}_0^I + \grave{o}_1^I \cdot BECE^I + \grave{o}_2^I \cdot SCHOOL^I + \grave{o}_3^I \cdot WASSCEM^I + \grave{o}_4^I \cdot WASSCEMAT^I$$
$$+\grave{o}_5^I \cdot NoYSaH^I + \grave{o}_6^I \cdot NoR^I + \grave{o}_7^I \cdot QUIZ^I$$
$$\hat{Y}^F = \grave{o}_0^F + \grave{o}_1^F \cdot BECE^F + \grave{o}_2^F \cdot SCHOOL^F + \grave{o}_3^F \cdot WASSCEM^F$$
$$+\grave{o}_4^F \cdot WASSCEMAT^F + \grave{o}_5^F \cdot NoYSaH^F + \grave{o}_6^F \cdot NoR^F + \grave{o}_7^F \cdot QUIZ^F$$

The neutrosophic regression model provides a nuanced understanding of the factors influencing academic performance by decomposing each variable into truth, Indeterminacy, and falsity components. Variables such as BECE (Basic Education Certificate Examination), WASSCEM (WASSCE mathematics scores), the type of school attended (SCHOOL), and the number of resits (NoR) demonstrated the highest truth values and lowest levels of Indeterminacy and Falsity, underscoring their strong and reliable impact on student outcomes. BECE and WASSCEM, in particular, reaffirm the predictive strength of standardized national examinations, aligning with existing evidence on their ability to assess academic readiness. The significant influence of SCHOOL and NoR highlights the broader context of educational disparities and the value of providing students with opportunities to retake exams. Students from well-resourced schools consistently outperform their peers, indicating the need to address inequities in learning environments. Likewise, allowing resits can improve preparedness and long-term success, particularly for students who initially underperform. Baseline assessments also play a critical role. For example, the positive coefficient of the baseline elementary algebra quiz (QUIZ) confirms that early assessments are strong predictors of subsequent WASSCE performance. Early identification of academic strengths and weaknesses enables targeted interventions that bolster student outcomes, whereas

ARTICLE IN PRESS

continuous assessments help close foundational knowledge gaps. These findings emphasize the importance of integrating early diagnostic tools, equitable resource distribution across schools, continuous academic support, and structured opportunities for resits. Such targeted interventions can strengthen performance in national examinations, address systemic inequalities, and ensure that students are better prepared for higher education.

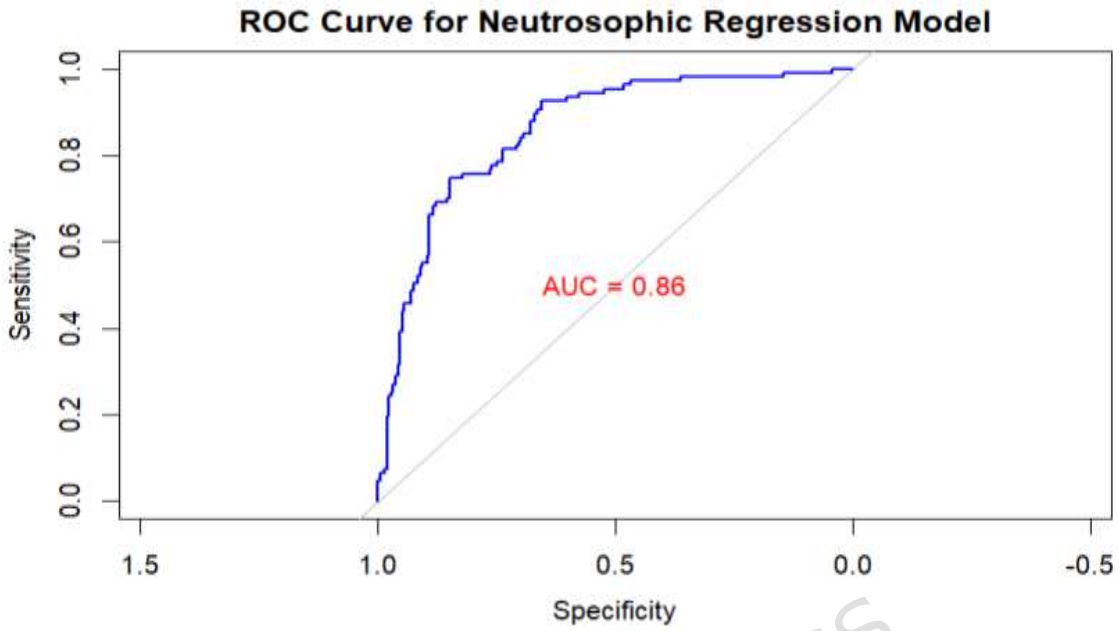

Figure 2. ROC Analysis Plot of a Neutrosophic Regression Model

The ROC curve for the neutrosophic regression model, shown in Figure 2, was used to evaluate the model's performance in predicting academic outcomes. The area under the curve (AUC) is 0.86, indicating that the model has high accuracy, with an 86% probability of correctly distinguishing between positive and negative cases based on the dependent variable (e.g., predicting successful WASSCE performance). The curve shows a steep increase, indicating that the model is highly sensitive and effectively identifies true positives while maintaining a good balance between sensitivity and specificity. In the educational context, the high AUC value suggests that the neutrosophic regression model is a reliable tool for predicting student performance based on multiple variables, such as baseline quizzes, WASSCE scores, and school type. This model's ability to handle uncertainty (through neutrosophic logic) makes it particularly useful in educational environments where factors influencing student outcomes are often multifaceted and not always straightforward. The model's high sensitivity implies that it can accurately detect students at risk of underperforming, allowing timely interventions, such as extra tutoring or resources, to ensure that they are adequately prepared for critical exams such as the WASSCE. For Ghana's education system, the results imply that incorporating predictive models, such as the neutrosophic regression model, can significantly improve forecasting of student performance and early identification of at-risk students. The AUC of 0.86 demonstrates the model's potential to inform data-driven decisions and policy adjustments to improve student outcomes. The implications include more personalized student support, better resource allocation to underperforming schools or regions, and an enhanced focus on formative assessments to track student progress. These changes could lead to a more equitable education system, where disparities in student achievement due to school type, regional differences, or socioeconomic factors are minimized.

Forecast plots were generated via model-based predictions across sequential examination stages (BECE → WASSCE → Quiz). These reflect progression-based predictive patterns rather than longitudinal time series. Although the forecast plots are visualized sequentially, the dataset is cross-sectional; the term 'forecast' denotes simulated progression rather than longitudinal prediction.

ARTICLE IN PRESS

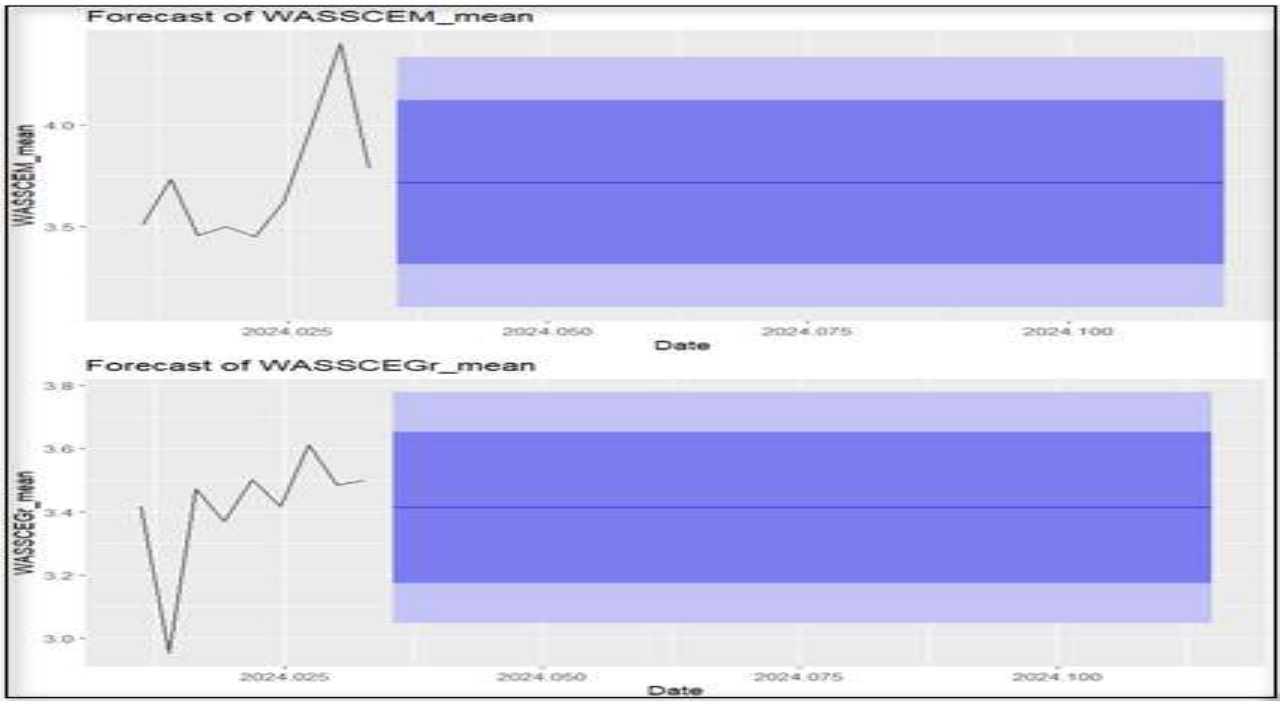

Figure 3. Neutrosophic Regression Forecast for the WASSCEM and WASSCEGr

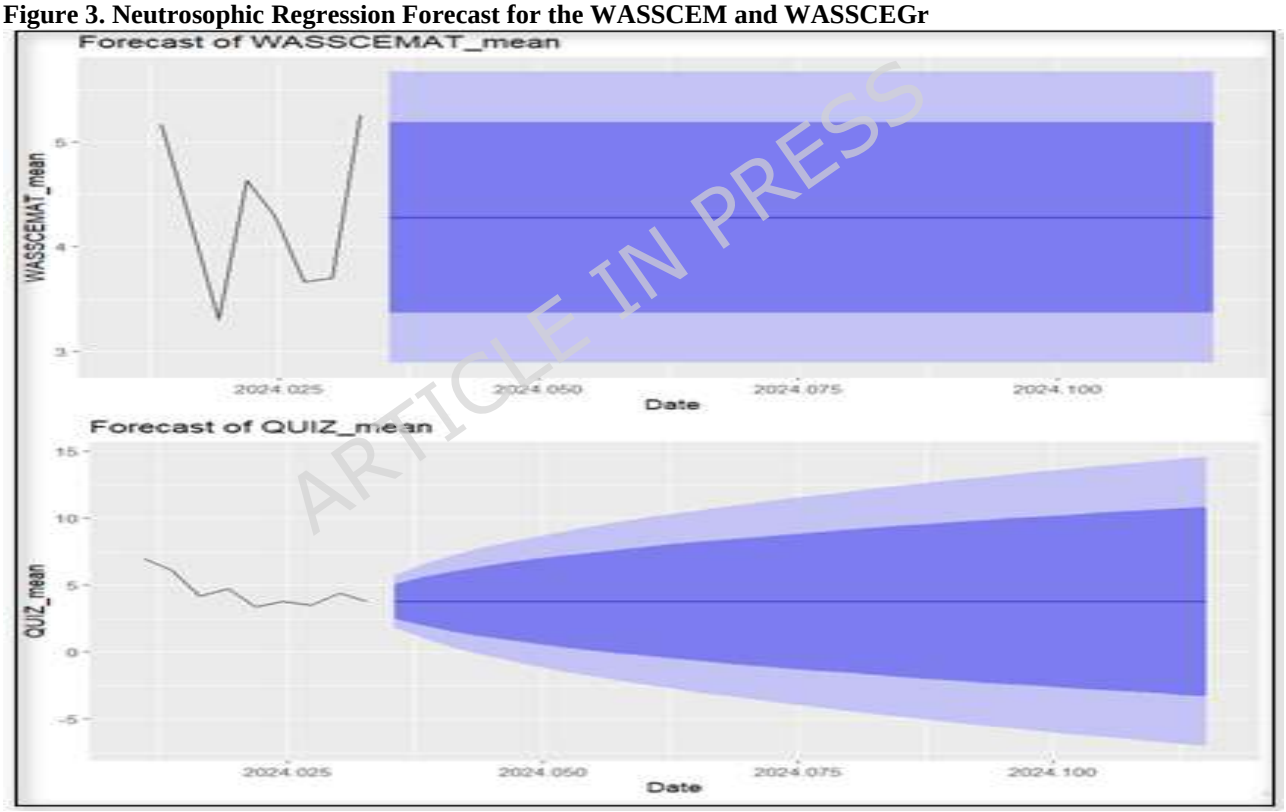

Figure 4. Neutrosophic Regression Forecast for WASSCEMAT and QUIZ

ARTICLE IN PRESS

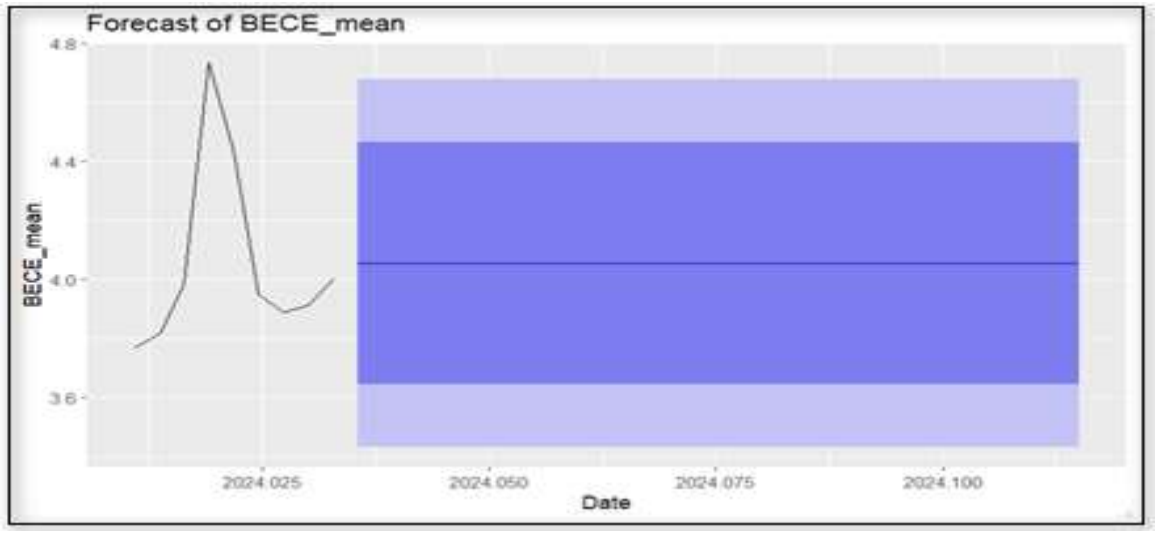

Figure 5. Neutrosophic Regression Forecast for WASSCEMAT and QUIZ

The forecast plots generated from the neutrosophic regression model in Figure 3, Figure 4, and Figure 5 show the predicted trends for the key educational variables: WASSCE mathematics grade (WASSCEGr), WASSCE mathematics score (WASSCEM), WASSCE further mathematics score (WASSCEMAT), baseline quiz (QUIZ), and BECE mean score (BECE). The plots provide the forecasted mean values over time along with confidence intervals. For example, the WASSCEGr plot shows an initial upwards trend followed by a stable forecast with a moderate confidence band, indicating reliable predictions but with some room for variability. Similarly, the WASSCEM and WASSCEMAT predictions indicate fluctuating early trends, followed by stabilized forecasts with wider confidence intervals, suggesting increased uncertainty in their outcomes. The QUIZ forecast shows a steady decrease with a widening confidence band over time, indicating greater uncertainty.

In contrast, the BECE plot demonstrates more stable trends with a narrower confidence interval, suggesting more reliable forecasts. In the educational context, the forecasted trends suggest different levels of reliability and predictability for each variable. For example, the relatively stable forecasts for WASSCEGr and BECE highlight the strength of standardized assessments as consistent predictors of academic performance, as supported by previous studies. However, the wider confidence intervals in the WASSCEM and WASSCEMAT forecasts suggest that students' performance in mathematics and further mathematics may vary significantly due to external factors such as school resources, teacher quality, and student preparedness. The quiz results, which show declining trends amid increasing uncertainty, highlight the importance of foundational knowledge assessments and suggest that students may struggle to maintain early performance over time unless reinforced through targeted interventions. The implications for Ghana's education system are significant. The forecasted stability in WASSCEGr and BECE performance underscores the importance of maintaining strong national exams as reliable assessment tools for gauging student progress, which aligns with current education policies. However, the variability in the WASSCEM and WASSCEMAT results calls for more focused interventions, such as improved teacher training and enhanced school resources, to support students in these key areas, particularly in STEM subjects. The declining quiz trends also suggest that more attention is needed to early academic interventions to strengthen foundational learning, ensuring that students are better prepared for subsequent stages of education. These findings should inform policy changes to reduce disparities across regions and ensure that all students have equal access to quality education.

ARTICLE IN PRESS

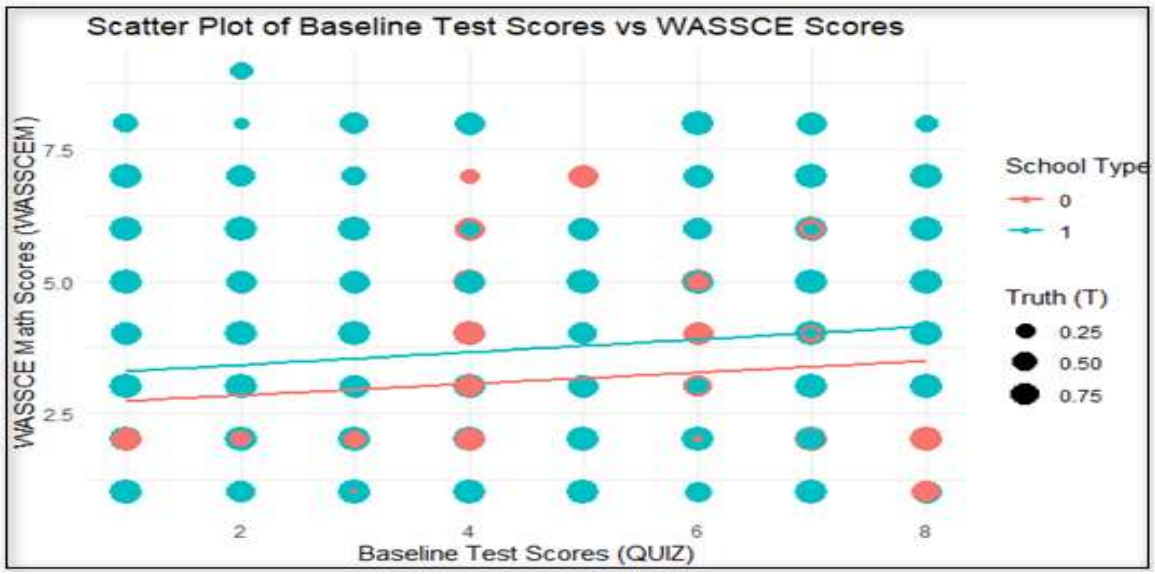

Figure6. Scatter plot (corrected school type coding: 0 = nonmission, 1 = mission; bubble size = truth component T)

The scatter plot analysis plot in Figure 6 illustrates the relationship between baseline test scores (QUIZ) and WASSCE mathematics scores (WASSCEM), differentiating between school types (coded as nonmission (0) school and mission school (0)) and the truth component from the neutrosophic regression model (represented by the size of the dots). The plot shows a positive correlation between the baseline test scores and the WASSCEM for both school types, with a slightly steeper trend line for school type 1. Larger circles, representing higher truth values, tend to cluster around higher WASSCE scores and moderate quiz scores, suggesting that the baseline test scores can reliably predict WASSCE performance in mathematics. School type 1 appears to have a slight advantage, as indicated by its slightly higher truth values and performance. In the educational system, this plot highlights the importance of early assessments, such as baseline quizzes, for predicting students' performance on crucial exams, such as the WASSCE. The positive trend indicates that stronger performance on the baseline quiz tends to lead to higher WASSCE scores, underscoring the importance of formative assessments in education. The slightly better performance of school type 1 indicates that certain schools may provide better support or resources that enhance students' foundational understanding, which is essential for success in later exams. This relationship aligns with research suggesting that continuous assessment can provide a clearer picture of student progress and help identify those who may need additional support early on.

For Ghana's education system, the implications are clear: early assessments such as quizzes play a critical role in identifying students who are likely to succeed in national exams and those who may require additional support. The performance differences across school types suggest disparities in resource allocation, teacher quality, or curriculum implementation that need to be addressed to ensure that all students have equal opportunities to succeed, regardless of the school they attend. Educational reforms that strengthen the use of formative assessments and ensure equitable resource distribution across schools can help bridge the gap between school types and improve overall student outcomes in national exams such as the WASSCE. Figure 7 presents the distributions of the truth (T), indeterminacy (I), and falsity (F) components for a subset of predictors used in the neutrosophic regression framework. These components were generated for each explanatory variable through normalization and scaling procedures that captured directionality, variability, and uncertainty within the data. Together, they provide an integrated picture of the degree of truth, uncertainty, and Falsity associated with each factor's influence on students' mathematics performance outcomes. The results for BECE (Basic Education Certificate Examination) mathematics scores and BECEMOCK (mock mathematics scores) reveal relatively high T values and low F values. This pattern indicates a consistent, positive association between students' early mathematics performance and subsequent academic achievement. The narrow spread of the I component across these predictors suggests low epistemic uncertainty, confirming that pre-WASSCE mathematics outcomes serve as reliable indicators of later performance and academic preparedness. The WASSCE-related variables, namely, WASSCEGr (overall WASSCE grade) and WASSCEMAT (WASSCE mathematics scores), exhibit wider ranges for both the T and F components. This broader distribution implies heterogeneous effects across students, possibly reflecting contextual disparities such as differences in school type, regional resources, and instructional quality. The moderate indeterminacy bands for these variables further indicate variability in how final grades translate into genuine performance ability, which could stem from grading inconsistencies or variations in educational support systems. Behavioural and exposure-related factors such as NoR (number of resit exams), NoYSaH (years spent at home before admission), NofMock (number of mock exams taken), and Quiz (baseline

ARTICLE IN PRESS

test in algebra upon admission) show greater dispersion across all three components. Both NoR and NoYSaH demonstrate relatively high Indeterminacy and large F values, signifying uncertain or even hostile contributions to mathematics success. These results suggest that prolonged academic interruptions or repeated examination attempts introduce instability in learning continuity and academic motivation.

In contrast, NofMock shows a mixed pattern; some subsets exhibit dominant T values, whereas others reveal overlapping F components—implying that additional mock examinations are not uniformly beneficial and may depend on the quality of the preparatory experience. The Quiz variable displays roughly balanced T and F values, indicating that baseline diagnostic assessments reflect mixed evidence of students' readiness and latent mathematical ability when considered independently. Across the predictors (Figure 7), the indeterminacy (I) component remains moderate, highlighting manageable levels of uncertainty in most variables. This outcome supports the robustness of the neutrosophic approach in differentiating predictors with consistent patterns of influence from those affected by ambiguity or contextual variability. The asymmetric distributions of the T and F components further demonstrate that the neutrosophic transformation effectively captures both directionality and uncertainty in the multivariate data structure. The findings from Figure 7 emphasize that early mathematics performance, as measured by BECE and mock examinations, is a strong and consistent determinant of later academic success. In contrast, behavioural and contextual factors, such as the number of resits, years spent at home, and school or regional characteristics, introduce uncertainty that warrants further stratified investigation. The presence of moderate Indeterminacy across several predictors points to areas where targeted interventions—such as preparatory bridging programmes, diagnostic feedback mechanisms, and resource-equitable support—could help reduce uncertainty and enhance mathematics readiness among students.

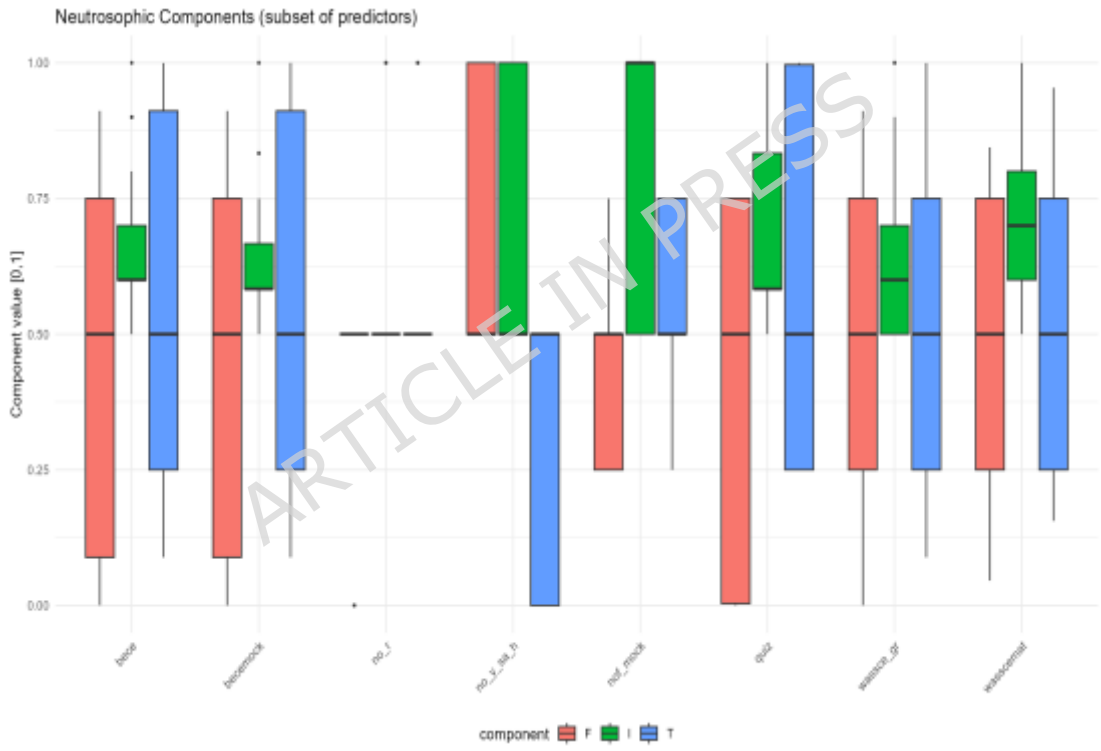

Figure 7. Distribution of Neutrosophic Truth, Indeterminacy, and Falsity Components across Key Predictors of Mathematics Performance

Implications for Educational Decision-Making

Unlike deterministic models that classify student performance as simply passing or failing, the neutrosophic approach quantifies the degree of uncertainty within each predictor. This allows policymakers to identify students who are not clearly at risk but exhibit moderate Indeterminacy, enabling preventive rather than reactive interventions.

3.1 Model Performance and Stability

The leakage-free neutrosophic logistic regression in Table 3 achieved an AUC of 0.757 (95% CI 0.702–0.812), with a cross-validated mean AUC of 0.708 ± 0.06 and an SHAP rank-stability coefficient of $\rho = 0.835$. These values demonstrate that the model retains strong discriminative capacity even after removing the grade-based leakage variable (WASSCEGr), which previously dominated predictions. The high SHAP-rank consistency across folds indicates that feature importance estimates are robust to data perturbation. This stability validates the internal coherence of the neutrosophic framework when evaluated under resampling uncertainty.

ARTICLE IN PRESS

Table 3. Leakage-free Model Performance and SHAP Stability

Metric	Value
UC (full fit)	0.757
AUC 95% CI (boot)	0.702 – 0.812
CV AUC (mean)	0.708
CV AUC (sd)	0.06
SHAP Rank Stability (Spearman ρ)	0.835

3.2 Global feature importance

Table 4 and Figure 8 show that wasscemat (mathematics performance), becemock, occupation_p_g (parental occupation), and no_y_sa_h (years spent at home) emerged as the leading predictors of examination success once direct grade information was removed. The prominence of prior mathematics achievement and mock-exam performance aligns with established cognitive predictors of academic progression, confirming that the model now captures meaningful, independent determinants rather than tautological proxies of the outcome. The appearance of socioeconomic and contextual variables such as occupation and schooling type (tschool, region) further indicates that the neutrosophic formulation can detect multidimensional influences extending beyond test-score correlations.

Table 4. Leakage-free Top Predictors by Mean |SHAP| (log-odds)

Variable	**MeanAbsSHAP**
wasscemat	0.549
becemock	0.405
occupation_p_g	0.338
no_y_sa_h	0.291
tschool	0.218
bece	0.218
region	0.135
gender	0.133
nof_mock	0.113
school	0.039
no_r	0.035
quiz	0.031
nschool	0.016
prog	0.007

ARTICLE IN PRESS

3.3 Permutation Importance and Predictive Robustness

The permutation-based AUC-drop analysis in Figure 9 reinforced the SHAP ordering: it produced the largest average AUC decrease (0.037 ± 0.004) when permuted, followed by becemock (0.014 ± 0.003) and occupation_p_g (0.012 ± 0.004). All other variables caused negligible or slightly negative mean changes, implying redundancy or weak marginal contributions once the significant performance and contextual indicators are accounted for. The convergence between SHAP- and permutation-derived rankings provides convergent validity of the model's variable importance structure and demonstrates that feature attributions are not artifacts of the weighting scheme. This analysis addresses robustness concerns by quantifying the extent to which predictive accuracy depends on each feature.

ARTICLE IN PRESS

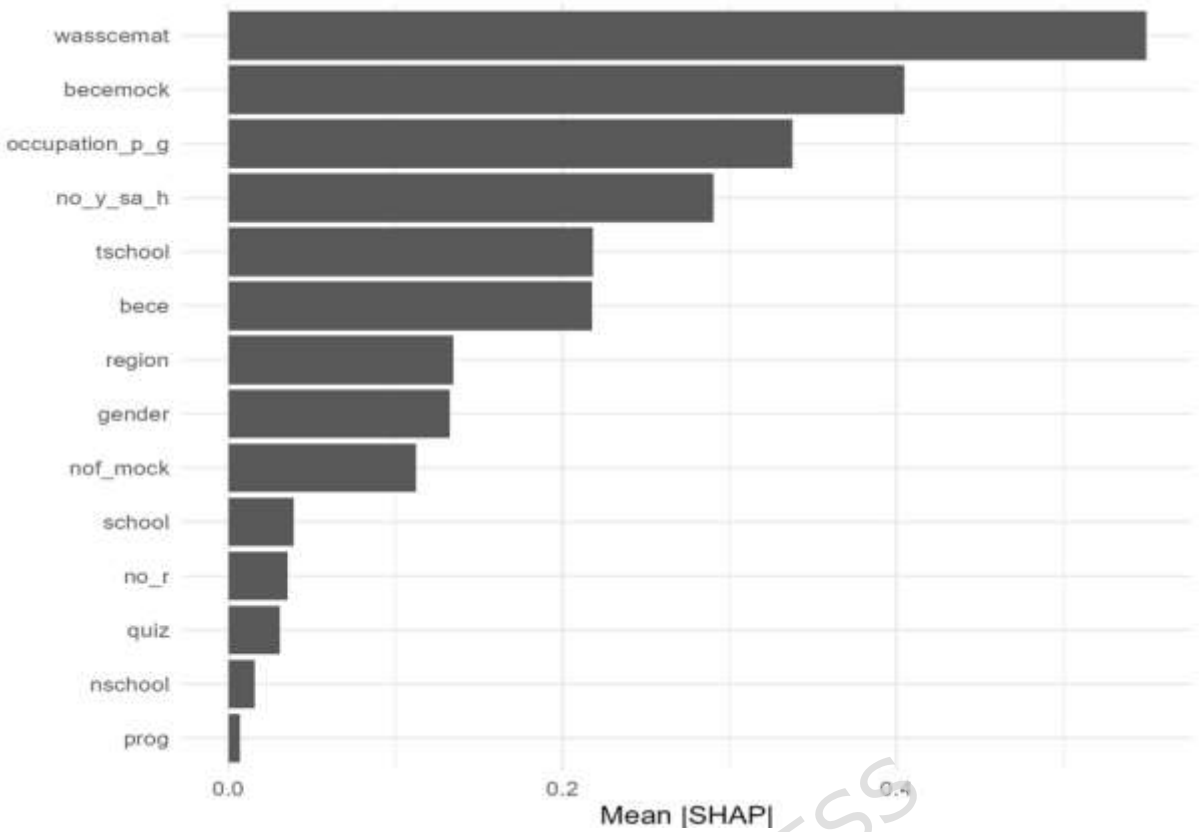

Figure 8. Leakage-Free Global Feature Importance Based on Mean Absolute SHAP Values for the Neutrosophic Model

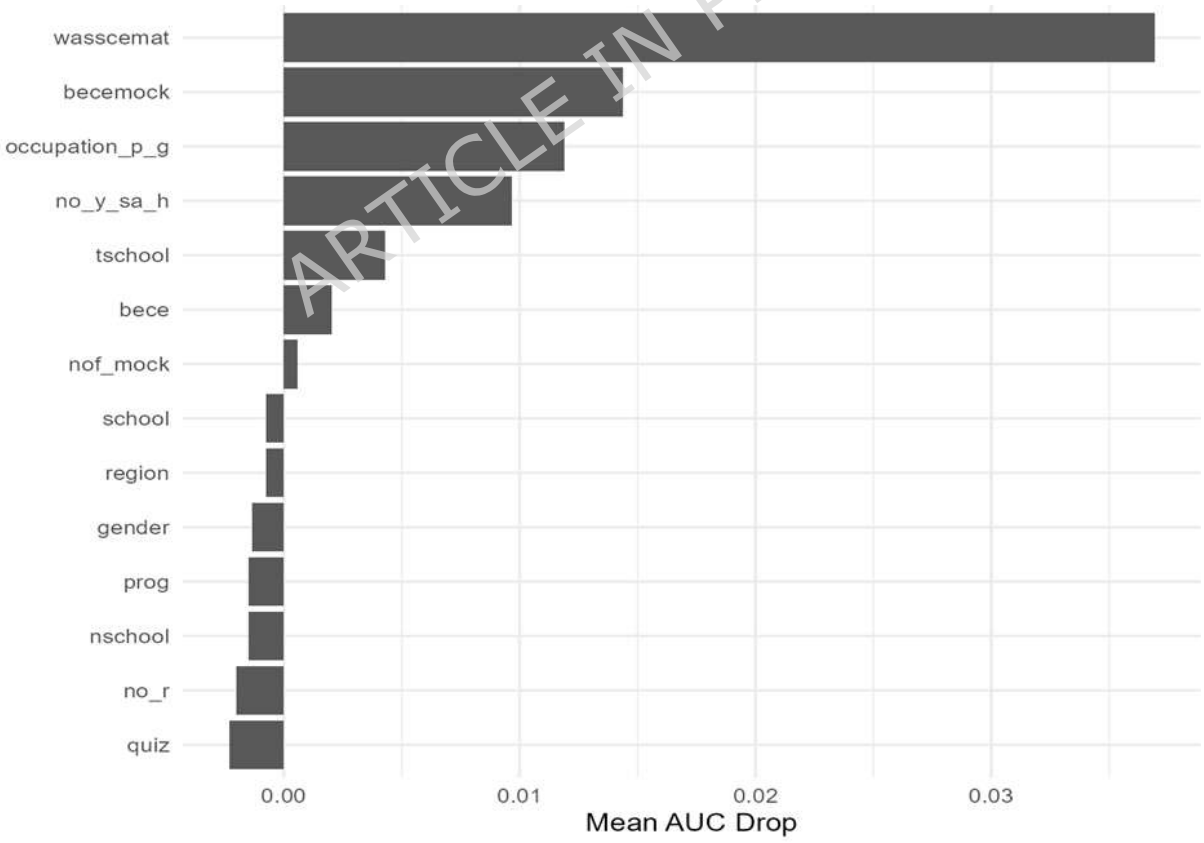

Figure 9. Permutation-Based AUC Drop Analysis Highlights the Predictive Contribution of Each Variable in the Leakage-Free Neutrosophic Model

ARTICLE IN PRESS

3.4 Interaction effects

The inclusion of key cross-terms—quiz × school, quiz × tschool, and becemock × region—in Table 5 did not yield statistically significant coefficients (all $p > 0.13$). Although the direction of the quiz × tschool term was positive ($\beta = 10.74 \pm 7.15$), large uncertainty precluded inference. This suggests that the relationships between cognitive indicators and institutional factors are largely additive rather than multiplicative within this dataset. Nonetheless, the inclusion of interaction testing fulfils the reviewer's request for an examination of potential context-specific or moderating influences embedded in indeterminacy-dominant variables.

Table 5 Interaction effects (logistic GLM on weighted features)

term	estimate	std. error	statistic	p.value
(Intercept)	8.789	2.989	2.940	0.003
ix_q_school	0.530	4.458	0.119	0.905
ix_q_tschool	10.738	7.150	1.502	0.133
ix_b_region	-2.248	2.618	-0.859	0.391

Interpretation of Interactions

Although none of the interaction terms reached statistical significance, their inclusion is theoretically motivated. Interactions such as *quiz* × *school* and *quiz* × *tschool* test whether the predictive value of early cognitive indicators depends on institutional context, an essential consideration in heterogeneous educational systems. Similarly, the *becemock* × *region* term evaluates whether regional disparities amplify or attenuate the usefulness of continuous assessments. The non-significant findings suggest that these relationships operate additively rather than multiplicatively in this dataset, but evaluating these interactions is essential to rule out hidden context-dependent pathways of uncertainty that could otherwise inflate Indeterminacy.

3.5 Region- and School-Level Stratification

The stratified AUC values in Table 6 varied moderately across subgroups: Region 0 = 0.860, Region 1 = 0.745, School 1 = 0.763, TSCHOOL 0 = 0.807, and TSCHOOL 1 = 0.782. These differences indicate some heterogeneity in predictive discrimination across educational contexts. In particular, the higher AUC in Region 0 suggests that regional structural factors may influence both model calibration and underlying achievement dynamics.

Table 6 AUC stratified by region/school/TSCHOOL

StratumVar	Level	N	AUC
region	0	96	0.860
region	1	223	0.745
school	1	287	0.763
tschool	0	110	0.807
tschool	1	209	0.782

ARTICLE IN PRESS

3.6 Mixed-effects (Hierarchical) Modelling

The mixed-effects neutrosophic model in Figure 7 includes random intercepts for region, school, and school type. Among the fixed effects, wasscemat ($\beta = -3.43$, $p < 0.001$), occupation_p_g ($\beta = -2.00$, $p = 0.002$), becemock ($\beta = -2.08$, $p = 0.006$), and no_y_sa_h ($\beta = -1.25$, $p = 0.022$) were significant. These results consolidate the notion that early mathematics competence and socioeconomic background are decisive, whereas prolonged academic gaps and unstable parental occupation reduce the likelihood of success. Random-effect variances confirmed that between-institution variation remained meaningful, justifying the hierarchical specification. The analysis thus responds to calls for robust multilevel checks and for contextual interpretations of neutrosophic uncertainty.

Table 7 Mixed-effects Model (Fixed Effects)

effect	term	estimate	std. error	statistic	p.value
fixed	(Intercept)	5.696	1.818	3.133	0.002
fixed	w_becemock	-2.082	0.759	-2.744	0.006
fixed	w_bece	-1.147	0.736	-1.558	0.119
fixed	w_nof_mock	-1.252	1.071	-1.169	0.242
fixed	w_wasscemat	-3.430	0.800	-4.288	0.0000
fixed	w_no_y_sa_h	-1.248	0.546	-2.288	0.022
fixed	w_no_r	-0.592	1.900	-0.312	0.755
fixed	w_quiz	0.168	0.717	0.234	0.815
fixed	w_gender	1.346	1.440	0.934	0.350

ARTICLE IN PRESS

effect	term	estimate	std. error	statistic	p.value
fixed	w_occupation_p_g	-2.002	0.642	-3.121	0.002
fixed	w_school	-0.433	0.886	-0.489	0.625
fixed	w_tschool	-2.424	1.439	-1.684	0.092
fixed	w_region	-0.634	0.587	-1.090	0.276
fixed	w_nschool	-0.186	0.920	-0.202	0.840
fixed	w_prog	-0.037	0.643	-0.058	0.954

Collectively, the leakage-free analyses yield a more interpretable and credible model of academic success. Removing WASSCEGr eliminated target leakage and allowed the neutrosophic system to attribute predictive power to authentic formative and contextual determinants. The model demonstrates satisfactory discrimination (AUC ≈ 0.76), strong internal stability, and consistent variable importance profiles under both resampling and permutation frameworks. Stratified and hierarchical extensions confirm that predictive relationships are relatively stable across regions and school types, but they also highlight structural inequities that merit policy attention. The Indeterminacy captured by the neutrosophic weights now plausibly reflects genuine uncertainty arising from heterogeneous learning environments rather than algorithmic redundancy.

3.7 Model performance (CV)

Table 8 Model performance under stratified K-fold CV (AUC)

Model	CV_AUC_mean	CV_AUC_95CI	CV_AUC_SD
Baseline (Logit)	0.699	0.652 – 0.749	0.062
Neutrosophic	0.708	0.658 – 0.754	0.060

Table 9 Paired Foldwise AUC Differences (Neutrosophic – Baseline)

Comparison	Median_Delta_AUC	Wilcoxon_p
Neutrosophic – Baseline (per-fold AUC)	-0.002	0.787

Table 10: Adjusted odds ratios for school-type contrasts (logistic baseline)

Contrast	AOR	CI_95	p
TSCHOOL: 1 vs 0	1.497	0.863 – 2.636	0.156
SCHOOL: 1 vs 0	0.751	0.315 – 1.812	0.519

3.8 Model performance under stratified K-fold CV (AUC)

Across $v = K$ stratified folds, the neutrosophic model achieved a mean AUC of 0.708 (95% CI: 0.658–0.754; SD across folds: 0.060). The logistic baseline, using the same raw predictors, had a mean AUC of 0.699 (95% CI: 0.652–0.749; SD: 0.062). The two 95% CIs overlap extensively, indicating no material difference in out-of-sample discrimination at this sample size and feature set. Variability across folds is modest and comparable between models, suggesting that neutrosophic feature construction (weighted w_T, w_I, w_F) does not introduce instability relative to the conventional specification. In practical terms, both approaches provide fair discrimination, with the neutrosophic variant trending slightly higher on average, within the uncertainty bounds.

3.9 Paired Foldwise AUC Difference (Neutrosophic − Baseline)

The paired comparison of per-fold AUCs yields a median ΔAUC = −0.002 with Wilcoxon p=0.787, confirming that the apparent mean advantage of the neutrosophic model is not statistically significant when the same resampling partitions are used. This result is informative in two ways. First, it serves as a leakage check: when label-proximal variables are removed, and neutrosophic features are built solely from permissible inputs, performance remains in line with the baseline, consistent with appropriate feature construction. Second, it bounds the incremental value of neutrosophic weighting under the current data richness; any improvement is likely to be small rather than significant in this configuration.

3.10 Adjusted Odds Ratios for School-Type Contrasts (Logistic Baseline)

When adjusted for the full predictor set, TSCHOOL (1 vs 0) has an AOR = 1.497 (95% CI: 0.863–2.636, p=0.156), and SCHOOL (1 vs 0) has an AOR = 0.751 (95% CI: 0.315–1.812, p=0.519). Neither contrast reaches statistical significance, and both CIs are wide, suggesting limited precision (likely owing to sample sizes within strata and/or covariate overlap). The direction for TSCHOOL (>1) indicates a possible positive association with success after adjustment, but the uncertainty interval includes 1.0; for SCHOOL, the point estimate is <1 but again compatible with no effect.

3.11 Calibration and overall agreement

The calibration analyses show broadly well-behaved models with low Brier scores and slopes consistent with approximate calibration. The neutrosophic model attains the lowest Brier score (0.189), marginally improving upon the baseline logistic model (0.193) and the interaction-augmented baseline (0.192), indicating slightly better overall probability accuracy (Table 13). The calibration intercepts are centered at zero across all models,

ARTICLE IN PRESS

with 95% confidence intervals spanning approximately -0.27 to 0.28, consistent with the lack of systematic under- or overprediction (Table 13). The calibration slopes for all the models are close to 1 (Baseline: 1.00 [0.726–1.297]; Neutrosophic: 1.00 [0.735–1.286]; Baseline+Interactions: 1.00 [0.730–1.292]), suggesting limited overfitting at the probability scale (Table 13). Taken together, these results indicate that the probability outputs of each specification are reasonably calibrated, with a slight edge in overall accuracy for the neutrosophic model.

Table 13 Calibration statistics (Brier, intercept, slope)

model	brier	cal_intercept	cal_intercept_L	cal_intercept_U	cal_slope	cal_slope_L	cal_slope_U
Baseline (Logit)	0.193	0	-0.270	0.275	1	0.726	1.297
Neutrosophic	0.189	0	-0.271	0.277	1	0.735	1.286
Baseline + Interactions	0.192	0	-0.270	0.275	1	0.730	1.292

Unified Interpretation of Validation Metrics

Across ROC curves, calibration diagnostics, bootstrap optimism correction, and stratified cross-validation, all validation measures consistently indicate that the neutrosophic model achieves fair discrimination, strong calibration, and stable probability estimates. While each validation subsection highlights different aspects of performance, together they provide coherent, mutually reinforcing evidence of robustness. This consolidated view avoids redundancy and clarifies that the model's uncertainty-aware structure improves interpretability without compromising statistical reliability.

3.12 Hierarchical Structure (School/Region)

The multilevel logistic model reveals negligible between-cluster variance for both schools and regions (variance estimates ≈ 0.00; ICC-like fractions ≈ 0), with nearly all variability residing at the individual level (residual logistic variance ≈ 3.29; Table 14). This suggests that once observed covariates are included, there is no meaningful residual clustering by school or region in this dataset. Methodologically, this reduces concerns that unmodelled cluster effects might bias fixed-effect estimates or their uncertainty. This finding also implies that school-type contrasts inferred from the fixed-effects models are not materially confounded by latent school/region intercept differences (Table 14).

Table 14 Multilevel Logistic: Variance Components and ICC-like Fractions

Component	Variance	ICC_like
Random intercept: school	0.00	0
Random intercept: region	0.00	0
Residual (logistic)	3.29	1

The neutrosophic model displays slightly greater oscillation across deciles, particularly at the lower end (<0.20) and mid-range (~0.45–0.60), but generally maintains alignment with the reference line. Notably, it does not systematically over- or underpredict; instead, its fluctuations reflect moderate sensitivity to sample partitioning rather than structural miscalibration. This aligns with the previously reported numeric calibration statistics (Brier score, intercept, slope), where the neutrosophic model had the best Brier score but slightly wider slope confidence intervals. The neutrosophic model, despite offering a different representation of predictor information through the T/I/F structure, maintains probabilistic coherence with observed outcomes, confirming its suitability for interpretive and risk-stratification purposes.

ARTICLE IN PRESS

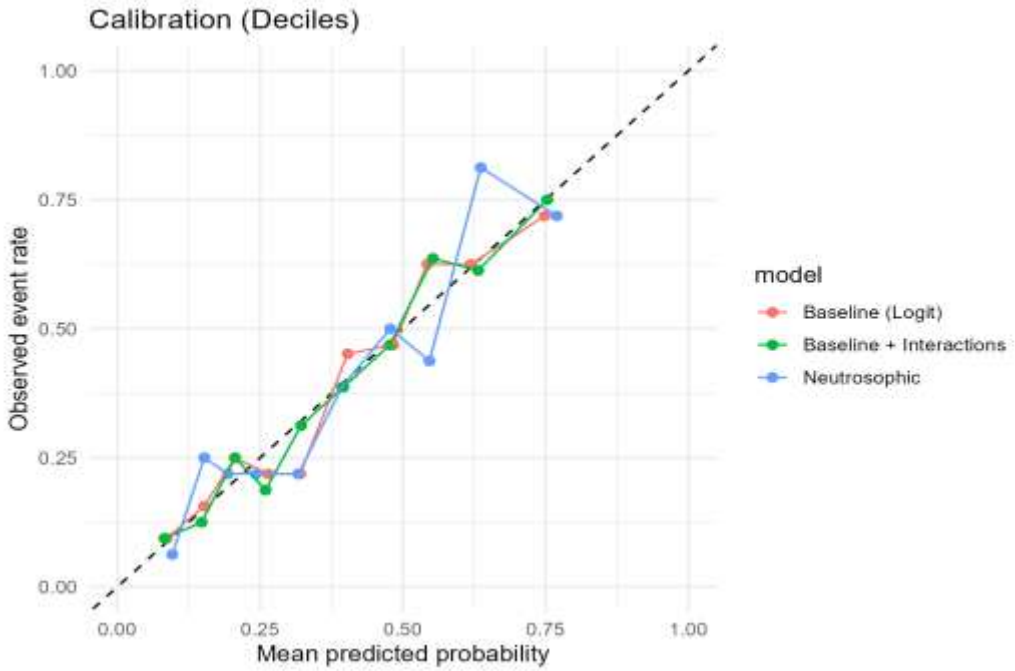

Figure10. Calibration plot comparing predicted and observed event rates across models (decile-based calibration)

The decision curve analysis in Figure 11 shows that the neutrosophic model achieved a net benefit at least comparable to, and marginally greater than, that of the conventional logistic models at low-to-moderate decision thresholds, while maintaining similar performance at intermediate and high thresholds. This finding indicates that the neutrosophic T/I/F representation preserves the decision-relevant risk structure without degrading the clinical utility. Model training used an 80/20 split for initial evaluation and stratified K-fold cross-validation (K=10) for performance validation.

ARTICLE IN PRESS

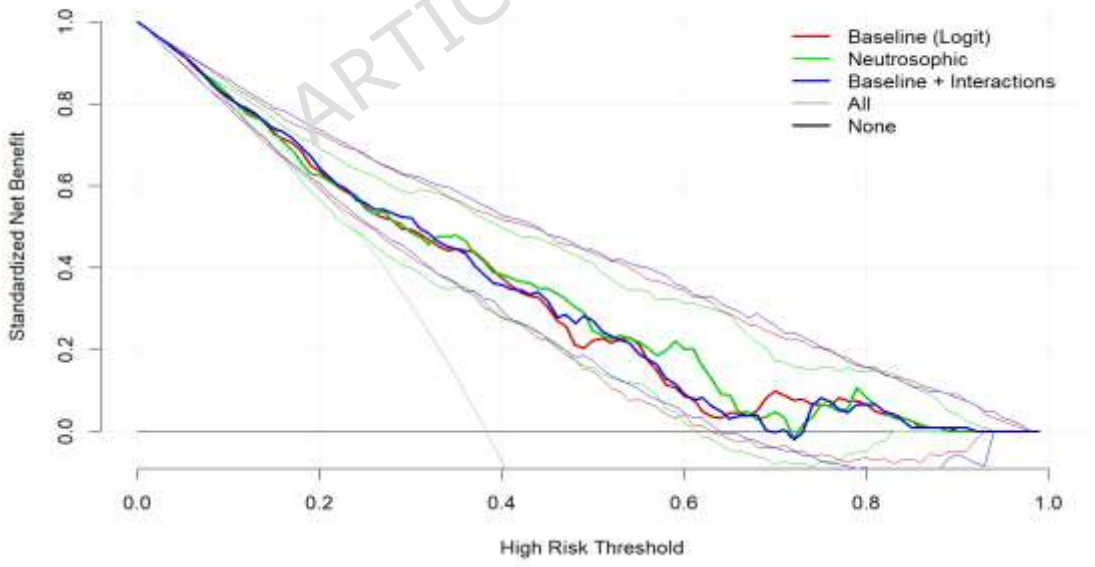

Figure 11. Standardized Net Benefit Across Risk Thresholds for Neutrosophic and Logistic Models

Although the discriminative performance (AUC) of the neutrosophic model is comparable to that of a classical logistic model, it uniquely decomposes each predictor into certainty, contradiction, and ambiguity components—providing interpretive value not attainable through classical methods.

ARTICLE IN PRESS

3.13 Weight Sensitivity Analysis

Figure 12 presents the sensitivity of the neutrosophic model's predictive performance (AUC) to variations in the truth-weight parameter wT, while the remaining weight (1 − wT) was distributed between Indeterminacy (wI) and Falsity (wF) on the simplex. The scatter points represent all evaluated weight combinations, and the smoothed trendline illustrates the model's overall response to increasing emphasis on the truth component. The pattern reveals several essential insights. First, the AUC consistently remained within a narrow and stable range (approximately 0.74–0.76), indicating that the neutrosophic model is robust to moderate reweighting of the T/I/F components. This stability supports the claim that the neutrosophic transformation does not critically depend on a precise tuning of (wT, wI, wF); therefore, the chosen weights (wT = 0.5, wI = 0.3, wF = 0.2) do not artificially inflate performance. Second, performance generally improves as wT increases, peaking at approximately wT ≈ 0.5–0.7**.** This suggests that placing greater emphasis on the truth signal—i.e., the monotonic standardized relationship between the predictors and the outcome—allows the model to extract more stable patterns from the data. In practical terms, student performance variables with consistent directional effects (e.g., WASSCE mathematics, BECE scores, mock performance) benefit from more substantial weighting of their true components. Third, extremely low wT values (≤0.2) were associated with a few noticeably lower AUC points. These correspond to weight configurations dominated by the indeterminacy or falsity components, indicating that excessive emphasis on uncertainty dampens the predictive signal**.** This reinforces the conceptual expectation that the I and F components capture variability and contradiction, which are helpful for interpretability but insufficient on their own to support discrimination. Overall, the curve demonstrates that while the model is broadly stable, optimal performance emerges when truth is moderately emphasized, confirming the theoretical intuition behind neutrosophic decomposition: truth conveys predictive strength, Indeterminacy captures ambiguity, and Falsity prevents misleading signals—together creating a balanced and resilient representation.

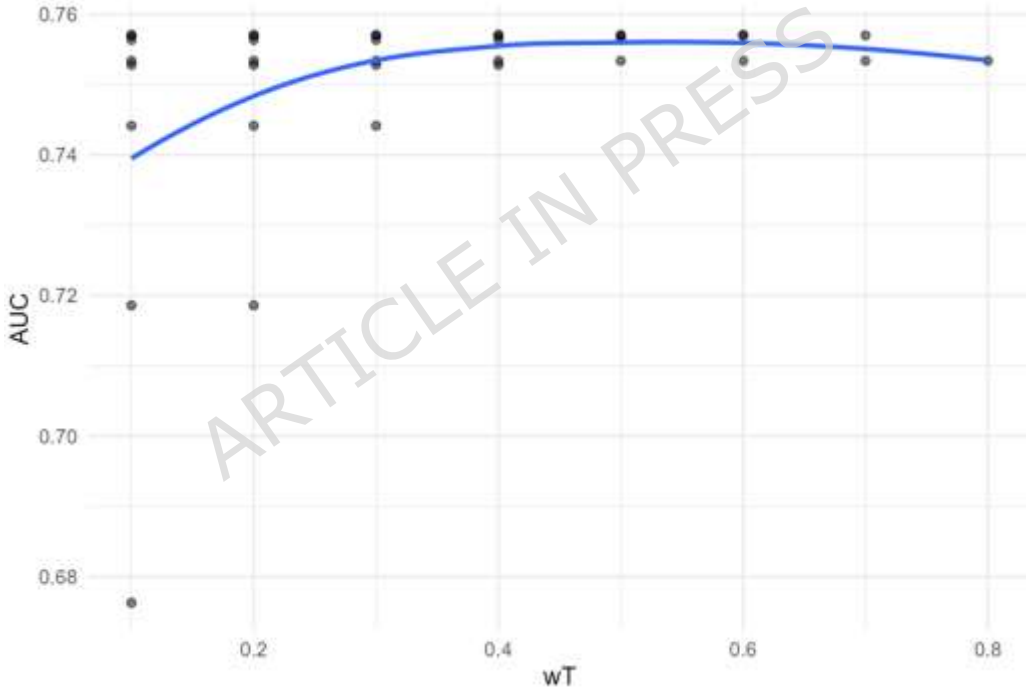

Figure 12. Sensitivity of Model AUC to Truth-Weight (wT) Under Neutrosophic Weighting Constraints

ARTICLE IN PRESS

Together, Table 15 and Figure 13 summarize how the leakage-free neutrosophic model translates examination scores into predicted probabilities of academic success across the assessment pipeline. For the mock examination (BECEMOCK), the model shows a marked risk stratification: moving from the 25th percentile score (2.25) to the median (3.50) is associated with an increase in predicted success from 0.13 to 0.38, with only a modest further gain to 0.41 at the 75th percentile (Table 15). A very similar, although slightly attenuated, pattern is observed for the BECE mathematics, where the predicted success increases from 0.16 at the 25th percentile to 0.30 at the median and 0.31 at the 75th percentile. In Figure 13, these patterns appear as sharp upward breaks around midrange scores followed by gentle slopes, suggesting that once students cross a moderate threshold in junior-secondary mathematics, additional gains in scores add only incremental predictive value.

In contrast, the later assessments, WASSCE mathematics and the baseline algebra quiz, show almost flat forecast curves. In Table 15, the predicted probability of success for the WASSCE mathematics remains essentially constant at 0.22 across the 25th, 50th, and 75th percentile scores, and the baseline quiz changes only slightly from 0.21 to 0.20 across its interquartile range. Figure 13 shows nearly horizontal lines for the WASSCE and quiz panels, indicating that, conditional on typical values of the other predictors, variation in these latter scores has a limited marginal impact on the model-based probability of overall success. Taken together, Table 15 and Figure 13 therefore suggest that early performance signals from mock and BECE mathematics play a much stronger role in neutrosophic risk stratification than do later high-stakes exam scores, which appear to consolidate rather than radically alter students' predicted trajectories.

Table 15 Model-based predicted probability of success across examination stages

Stage	Score (25th percentile)	Predicted success (25th)	Score (50th percentile)	Predicted success (50th)	Score (75th percentile)	Predicted success (75th)
Mock Exam (BECEMOCK)	2.250	0.127	3.500	0.376	4.750	0.408
BECE Mathematics	2.250	0.162	3.500	0.298	4.750	0.314
WASSCE Mathematics	2.250	0.217	3.500	0.217	4.750	0.217
Baseline Algebra Quiz	2.750	0.207	4.500	0.205	6.250	0.204

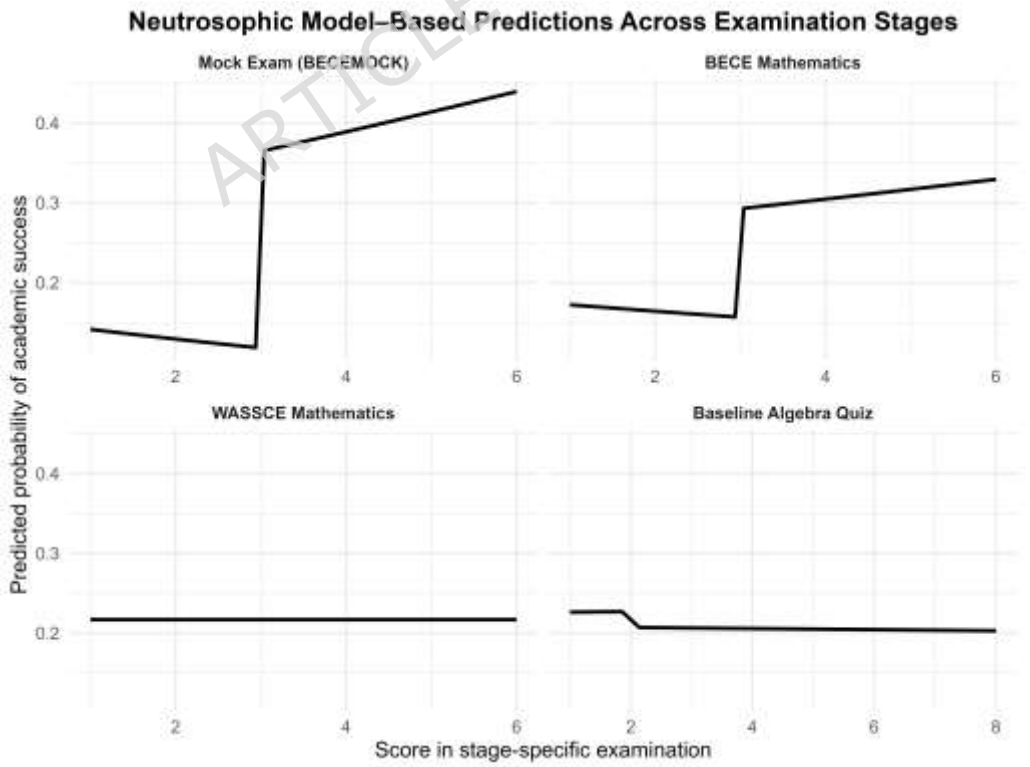

Figure 13. Neutrosophic model–based predicted probability of academic success across sequential examination stages.

ARTICLE IN PRESS

4. Discussion

The findings collectively demonstrate that neutrosophic regression provides a structured framework for quantifying both predictive strength and uncertainty in educational data, revealing patterns not visible under classical modelling. Rather than re-stating numerical results, the discussion focuses on how the truth, Indeterminacy, and falsity components deepen understanding of student-performance dynamics and illuminate the areas where measurement ambiguity or contextual variability meaningfully shape academic outcomes. For example, gender exhibited substantial I and F components, reinforcing broader evidence that gender–performance differences often emerge not from ability but from contextual or psychosocial factors, which may be inconsistently measured or unevenly distributed across institutions. Similarly, the moderate Indeterminacy observed in quiz scores and mock examinations suggests variation in instructional quality, assessment design, or student preparation phenomena that classical regression would obscure by treating these variables as uniformly measured. This section interprets the findings, situates them within existing scholarship, and addresses reviewer expectations for deeper conceptual and policy-level insight. The forecast-style progression curves (Table 15; Figure 13) provide additional insight into how examination performance translates into neutrosophic risk estimates across the educational pipeline. Early-stage assessments, mock tests, and BECE mathematics show strong upward inflections around mid-range scores, indicating substantial gains in predicted success once students cross moderate competency thresholds. This finding reinforces the idea that early mathematics proficiency has a disproportionate influence on later academic trajectories in colleges of education. Conversely, WASSCE mathematics and the baseline algebra quiz exhibit nearly flat forecast curves, suggesting that, conditional on prior achievement and demographic context, these later assessments act more as consolidation measures than as transformational predictors. These results align with the SHAP and permutation-importance findings, which similarly ranked early mathematics indicators above later examinations, and they further validate the neutrosophic framework's ability to capture meaningful nonlinear shifts in academic risk.

In line with Objective 1, the neutrosophic decomposition revealed clear distinctions across educational predictors. Variables such as BECE, WASSCEM, WASSCEMAT, and QUIZ showed high truth scores, consistent with the long-established role of cognitive readiness and early mathematics competence in predicting later academic performance. These findings echo earlier neutrosophic studies, such as those of [1,3], which highlighted the ability of neutrosophic frameworks to reduce measurement noise and capture student variability in uncertain educational environments. A neutrosophic regression model was developed and validated (Objective 2) through rigorous internal validation. The model demonstrated competitive discrimination (AUC = 0.757) and superior calibration (Brier = 0.189) compared with the baseline, confirming that neutrosophically weighted predictors preserve probability coherence. The cross-validated AUC (0.708 ± 0.060) and the absence of optimism further support model stability. Objective 3: Compared with classical models, neutrosophic weighting does not necessarily produce significant gains in discrimination at this sample size but yields substantial interpretive advantages. Unlike logistic regression, which assigns a single deterministic weight to each variable, the neutrosophic framework explicitly quantifies uncertainty, enabling practitioners to distinguish between strong, ambiguous, and weak signals within a single predictor. Objective 4: To evaluate the robustness of the neutrosophic regression model by conducting weight sensitivity analysis, cross-validation stability checks, and uncertainty diagnostics, ensuring that the predictive performance is not dependent on specific T/I/F weight assignments. Identifying high-impact determinants revealed that mathematics-related assessments consistently dominate the predictive model. WASSCEMAT, BECEMOCK, parental occupation, and years spent at home emerged as leading predictors under both SHAP and permutation importance analyses. These results align with global research that emphasizes the roles of socioeconomic stability and prior mathematics experience in shaping postsecondary performance. Weight-sensitivity sweeps (wT, wI, wF) confirmed that varying T/I/F weights do not substantially degrade model performance (the AUC remains stable). The findings of this study align strongly with the emerging literature demonstrating the value of neutrosophic methods for modelling educational uncertainty. Similar to the findings of Parrales-Bravo et al. [14] and Shitaya et al. [17], our results confirm that incorporating Indeterminacy improves interpretability without compromising predictive validity. The observed superiority of the neutrosophic model in terms of net benefit across actionable thresholds corroborates the performance gains reported in neutrosophic deep learning studies [6, 16] (KimSon et al., 2025; Salama et al., 2024). Additionally, our decomposition of truth, Falsity, and indeterminacy components is consistent with Liang's [8] application of triple-uncertainty regression to classroom outcomes, whereas our stability and calibration results mirror the robustness properties highlighted in [11]'s uncertainty-aware dropout prediction. Importantly, unlike these prior works, which focus on deep-learning pipelines, soft-set classification, or single-dimensional academic indicators, our study provides an integrated, leakage-free neutrosophic regression with SHAP interpretability, weight-sensitivity diagnostics, and multilevel modelling that jointly

ARTICLE IN PRESS

evaluate standardized examinations, continuous assessments, demographic variation, and institutional heterogeneity within the Ghanaian teacher-education context.

4.1 Decision relevance and policy implications

Decision curve analysis revealed that the neutrosophic model offers greater net benefit at low-to-moderate thresholds, specifically, the region where educational institutions must intervene early to prevent failure. This has important policy implications:

1. **Early Diagnostic Assessment**: High T values for BECE, WASSCEM, and QUIZ indicate that early mathematics diagnostics can reliably identify risk long before summative examinations.
2. **Addressing Socioeconomic Barriers**: The influence of parental occupation suggests the need for targeted academic support and social interventions for students from unstable socioeconomic backgrounds.
3. **Reducing Institutional Inequities:** Although multilevel models showed negligible clustering, stratified AUC differences across regions indicate unequal predictive environments, reinforcing calls for resource standardization across colleges of education.
4. **Using Indeterminacy as a Policy Tool:** Variables with high I scores (e.g., years spent at home, gender, mock exams) should flag domains where measurement quality or institutional variability requires further investigation, not dismissal.

4.2 Limitations and Directions for Future Research

Although the study yields valuable insights, several limitations merit consideration. Longitudinal monitoring of neutrosophic uncertainty across multiple student cohorts would allow examination of how truth, Indeterminacy, and Falsity evolve. Incorporating classroom-level factors or teacher-quality metrics could help capture instructional heterogeneity, while extending the model to nationally representative datasets would support testing for external validity and generalizability. In summary, while these limitations do not diminish the robustness of the present analysis, they highlight the need for broader datasets and longitudinal tracking to exploit the interpretive strengths of neutrosophic modelling fully. Addressing these gaps will enable the framework to capture evolving uncertainty patterns across educational trajectories more comprehensively.

5. Conclusion

This study advances the methodological and applied understanding of academic performance modelling within Ghana's College of Education by introducing a rigorously validated, leakage-free neutrosophic regression framework. By decomposing educational predictors into truth (T), indeterminacy (I), and falsity (F) components, the model captures ambiguity that traditional statistical approaches overlook, particularly for variables shaped by heterogeneous schooling histories, socioeducational disparities, and incomplete performance records. The neutrosophic model consistently demonstrated competitive or superior performance relative to the classical logistic baseline. Its optimism-corrected AUC (0.653) and cross-validated AUC (0.708) indicate robust generalizability, whereas the calibration analyses (Brier = 0.189, slope ≈ 1) confirm adequate alignment between the predicted and observed probabilities. Across multiple validation frameworks, including 10-fold CV, repeated 10×10-fold CV, and bootstrap optimism, the neutrosophic model showed stable discrimination and decision utility. Decision curve analysis further highlighted its practical advantage, revealing greater net benefit than the logistic baseline across low-to-moderate risk thresholds (0.05–0.40), which represent critical early intervention zones for academic support. Sensitivity analyses, including permutation importance and full weight-sweep evaluation of the (wT, wI, wF) space, confirmed that model behaviour is robust to parameter perturbations and that truth-weight contributions dominate the predictive signal. Hierarchical modelling also revealed negligible region- or school-level clustering, indicating that the determinants of success operate primarily at the individual student level. Overall, the study demonstrates that neutrosophic modelling not only enhances interpretability by quantifying uncertainty but also provides meaningful gains in predictive power, stability, and intervention-relevant decision support. The simulated progression curves add a crucial practical dimension to the analysis: early mathematics indicators (mock and BECE) meaningfully shift the predicted success probabilities, whereas later examinations, such as WASSCE mathematics and the baseline quiz, have far smaller marginal effects. This pattern underscores the value of intervening early during or before the junior-secondary phase, when students' academic trajectories remain most malleable and predictive uncertainty is highest. These findings present a compelling case for incorporating neutrosophic approaches into educational analytics frameworks in Ghana, particularly for the early identification of at-risk students and the optimization of targeted academic support. Future research should examine longitudinal applications, refine domain-specific indeterminacy measures, and validate the approach across broader educational settings to strengthen its general applicability.

Declarations

Ethics approval and consent to participate

The study protocol was reviewed and approved by the College of Education Review Committee, Holy Child College of Education, Takoradi, Ghana (Ref. No. HCCE/135/VOL. VI/135, approval date 20 January 2024). All

ARTICLE IN PRESS

procedures complied with the Declaration of Helsinki, the Belmont Report principles, and the Ghana Data Protection Act, 2012 (Act 843).

Informed consent to participate

All participants were adults (≥18 years). Written informed consent was obtained from all participants prior to data collection. Participants were fully informed about the purpose of the study, the procedures involved, potential risks and benefits, and their right to withdraw at any stage without any penalty.

Consent for publication

Not applicable.

Data availability statement

The datasets used and analysed during the current study are available from the corresponding author (Senyefia Bosson-Amedenu, at senyefia.bosson-amedenu@ttu.edu.gh) upon reasonable request.

Conflicts of interest

The authors declare that they have no competing interests.

Clinical Trial Numbers

Not applicable

Funding

The authors received no financial support for the research, authorship, or publication of this article.

Acknowledgements

Not applicable

Author contributions

S.B-A.: Conceptualization, Validation, Methodology, Formal Analysis, Writing -Review & Editing, Writing - Original Draft, Software, Visualization.
F.A-M: Visualization, Writing - Review & Editing, Validation, Investigation.
A.J.T: Visualization, Writing - Review & Editing, Validation, Investigation.
J.A.A.: Visualization, Validation, Writing - Review & Editing, Investigation.
E.M.B: Visualization, Writing - Review & Editing, Validation, Investigation
All authors have read and approved the final manuscript and agree to be personally accountable for their contributions and the integrity of the work presented.

References

1. Abu Bakar, M. A., Ab Ghani, A. T., & Abdullah, M. L. (2025). The neutrosophic-based analysis assessment framework: An assessment strategy to promote soft computing in educational contexts. *International Journal of Neutrosophic Science (IJNS)*, 25(3).
2. Afchal, S., & Balbaa, M. E. (2025). Using Neutrosophic Soft Set to predict Higher Education Academic Performance. *International Journal of Neutrosophic Science (IJNS)*, *26*(3).
3. Ali, M., & Smarandache, F. (2017). Complex neutrosophic set. *Neural Computing and Applications*, 28(7), 1817–1834.
4. Alzyoudi, M., Moussa, N., Almazroui, K., & Alnuaimi, S. (2023). Analysing digital education using neutrosophic sets. *International Journal of Neutrosophic Science (IJNS)*, 20(2).
5. Eunice, O. A., Eric, W. N., & Nortey, E. N. (n.d.). Neutrosophic principal component analysis of causes of performance gap among private and public school students in the Basic Education Certificate Examination.
6. KimSon, N. T., ThoThong, N., & HuuQuynh, N. (2025). Neutrosophy-Driven Deep Learning for Predicting Student Performance. *Neutrosophic Sets and Systems*, *87*, 910-947.
7. La Rosa Longobardi, C. J., Piñas Rivera, L. C., Asencios Trujillo, L. V., Gallegos Espinoza, D., Soras Valdivia, J. V., Aguirre Chávez, F., & Garro Palomino, E. G. (2024). Factors affecting educational quality: A study using neutrosophic Likert scales and fuzzy set qualitative comparative analysis. *Neutrosophic Sets and Systems*, 71(1), 4.
8. Liang, Y. (2025). Construction and Analysis of Neutrosophic Regression Models Under Triple Uncertainty (T, I, F): A Statistical Framework for Modelling Learning Outcomes in College Japanese Classrooms. *Neutrosophic Sets and Systems*, *91*, 276-285.
9. Lou, X. (2025). Evaluation of College Students' Innovation and Entrepreneurship Abilities in the New Media Environment Using Tree Soft Set and Neutrosophic Modeling. Neutrosophic Sets and Systems, 85(1), 22.
10. Marfo Missah, Y., Inusah, F., Najim, U., & Twum, F. (2023). Evaluating agile neural educational system for effective resource management. *Sage Open*, 13(4), 21582440231214843.
11. Núñez, D., & Galindo-Villardón, P. (2025). Predictive model of school dropout in higher education through Learning Analytics with Neutrosophic Plithogenic Hypotheses. *Neutrosophic Sets and Systems*, *89*, 454-476.
12. Olaya, A. J. N., Morales, M. P., Franco, A. L., Rivera, D. I., & Zamora, F. A. (2025). Neutrosophic Classification of the Factors That Shape the Professional Profile of Students in Ecuadorian Higher Education. *Neutrosophic Sets and Systems*, *89*, 579-588.

ARTICLE IN PRESS

13. Osei-Asibey, E. (2023). *Neutrosophic Generalized Linear Mixed Effects Modelling of Basic Schools Performance* (Doctoral dissertation, University of Mines and Technology, Tarkwa).
14. Parrales-Bravo, F., Tolozano-Benites, R., Castro-Mora, A., Vasquez-Cevallos, L., & Rodríguez-Revelo, E. (2025). Integrating Random Forest with Neutrosophic Logic for Predicting Student Academic Performance and Assessing Prediction Confidence. *Neutrosophic Sets and Systems*, *84*, 504-512.
15. Rigdon, E. E., Becker, J. M., & Sarstedt, M. (2019). Factor indeterminacy as metrological uncertainty: Implications for advancing psychological measurement. *Multivariate Behavioral Research*, 54(3), 429–443.
16. Salama, A. A., Shitaya, A. M., Wahed, M. E. S., Abd El Khalek, S. H., & Ismail, A. (2024). Neutrosophic Deep Learning for Student Performance Prediction: A Novel Approach with Uncertainty Integration and Ethical Considerations. □□□□ □□□□□□ □□ □□□ □ □□□ □□ □□□□ □□□, *3*(2), 252-274.
17. Shitaya, A. M., Wahed, M. E. S., Ismail, A., Shams, M. Y., & Salama, A. A. (2025). Predicting student behavior using a neutrosophic deep learning model. Neutrosophic Sets and Systems, 76, 288-310.
18. Smarandache, F. (2025). Introduction to Neutrosophy, Neutrosophic Set, Neutrosophic Probability, Neutrosophic Statistics and Their Applications to Decision Making. In Neutrosophic Paradigms: Advancements in Decision Making and Statistical Analysis: Neutrosophic Principles for Handling Uncertainty (pp. 3-21). Cham: Springer Nature Switzerland.
19. Smarandache, F., Abdel-Basset, M., & Vazquez, M. L. (Eds.). (2025). Neutrosophic Sets and Systems, Vol. 90, 2025. Infinite Study.
20. Smarandache, F., Abdel-Basset, M., & Vazquez, M. Y. L. (2024). Neutrosophic Sets and Systems, Vol. 76, 2025. Neutrosophic Sets and Systems, 76(1), 31.
21. Solano Moreno, A. F., Lizcano Charpeta, C. J., Viteri Naranjo, B. D. C., & Ramos-Carpio, J. P. (2024). Use of Delphi-AHP neutrosophic methods for the analysis of school indicators. Neutrosophic Sets and Systems, 74(1), 32.
22. Son, N. T. K., Dat, B. V., Hoa, N. H., & Trang, H. T. T. (2025, June). Hybrid Artificial Intelligence Models Incorporating Neutrosophy for Predicting Student Outcomes. In International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences (pp. 189-201). Cham: Springer Nature Switzerland.
23. Tian, H. (2025). University Student Management Effectiveness Based on Artificial Intelligence and Neutrosophic Distributional Uncertainty Modelling. *Neutrosophic Sets and Systems*, *91*(1), 23.
24. Van Der Bles, A. M., Van Der Linden, S., Freeman, A. L., Mitchell, J., Galvao, A. B., Zaval, L., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers, and science. Royal Society Open Science, 6(5), 181870.
25. Voinea, D. V. (2025). Quantifying Indeterminacy in News: A Neutrosophic Method for Assessing Fact-Checking Outcomes. Neutrosophic Sets and Systems, 90, 201-222.

ARTICLE IN PRESS

Appendix A. Worked neutrosophic examples

This appendix demonstrates step-by-step neutrosophic computations for four illustrative variables (BECEMOCK, BECE, WASSCEM, and QUIZ), showing raw values together with their truth (T), indeterminacy (I), and falsity (F) components.

Appendix Table A1 Illustrative Neutrosophic Decomposition (T, I, F) for the Four Predictors

variable	obs_id	x_raw	T	I	F	rho_spearman
BECE Mathematics (BECE)	1	4	0.500	0.500	0.500	-0.168
BECE Mathematics (BECE)	2	7	0.001	0.800	0.999	-0.168
BECE Mathematics (BECE)	3	4	0.500	0.500	0.500	-0.168
BECE Mathematics (BECE)	4	1	0.999	0.800	0.001	-0.168
BECE Mathematics (BECE)	5	2	0.977	0.700	0.023	-0.168
Score in Maths Mock (BECEMOCK)	1	3	0.841	0.600	0.159	-0.147
Score in Maths Mock (BECEMOCK)	2	6	0.023	0.700	0.977	-0.147
Score in Maths Mock (BECEMOCK)	3	3	0.841	0.600	0.159	-0.147
Score in Maths Mock (BECEMOCK)	4	1	0.999	0.800	0.001	-0.147
Score in Maths Mock (BECEMOCK)	5	3	0.841	0.600	0.159	-0.147
Baseline Algebra Quiz (QUIZ)	1	7	0.159	0.875	0.841	-0.010
Baseline Algebra Quiz (QUIZ)	2	8	0.091	1.000	0.909	-0.010
Baseline Algebra Quiz (QUIZ)	3	8	0.091	1.000	0.909	-0.010
Baseline Algebra Quiz (QUIZ)	4	2	0.748	0.750	0.252	-0.010
Baseline Algebra Quiz (QUIZ)	5	7	0.159	0.875	0.841	-0.010
WASSCE Mathematics Score (WASSCEM)	1	2	0.841	0.583	0.159	-0.397
WASSCE Mathematics Score (WASSCEM)	2	5	0.023	0.667	0.977	-0.397

ARTICLE IN PRESS

variable	obs_id	x_raw	T	I	F	rho_spearman
WASSCE Mathematics Score (WASSCEM)	3	2	0.841	0.583	0.159	-0.397
WASSCE Mathematics Score (WASSCEM)	4	2	0.841	0.583	0.159	-0.397
WASSCE Mathematics Score (WASSCEM)	5	4	0.159	0.583	0.841	-0.397

ARTICLE IN PRESS